

ON INTELLIGENCE

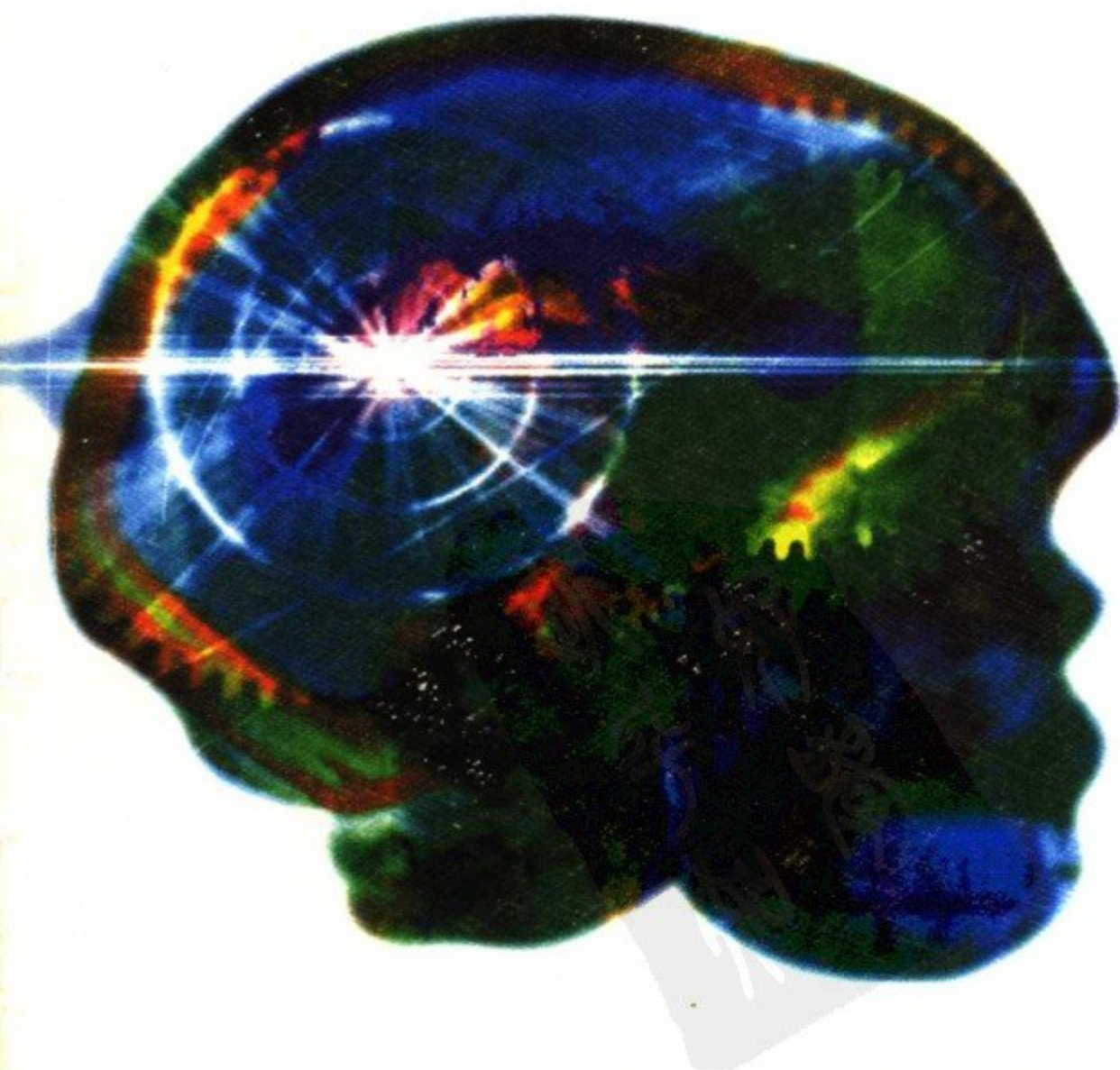
真正认识人类大脑是开发智能机器的必由之路！

—— 杰夫·霍金斯

人工智能的未来

[美] 杰夫·霍金斯 (Jeff Hawkins) 著
桑德拉·布拉克斯莉 (Sandra Blakeslee)

贺俊杰 李若子 杨倩 译



陕西科学技术出版社

ON INTELLIGENCE

发明PalmPilot的杰夫·霍金斯认为，他能让机器像人类一样创造性地工作。大话吧？可能是。但所有专家却不敢嘲笑。

——《商业周刊》

《人工智能的未来》是一部划时代的作品，它首次将人们期待已久的、有关人类大脑功能的重大理论诠释清楚。书中俯拾即是睿智的思想，是有关人类大脑功能原理的必读资料。

——美国科学院院士、加州大学神经病理学教授 MIKE MERZENICH

《人工智能的未来》一书势必影响深远，人人都应该读读它。1943年物理学家Erwin Schrodinger所著经典作品《生命是什么》，解开了当时的头号难题——分子是如何储存遗传基因密码的。而这本书是让众人真正认识人类大脑结构的。两书难分伯仲。

——美国Cold Spring Harbor 实验室负责人，1962年诺贝尔生理学奖得主 JAMES D. WATSON

杰夫·霍金斯所著的这本书，独特、新颖，振聋发聩，是一本可读性极强的作品，它对人类大脑皮层所具有的知觉、认识、行为和智能功能提出了新的理论构想。它的许多章节，尤其是讲述智能、创造力、硅晶大脑的部分，会令读者印象深刻、回味无穷。

——2000年诺贝尔生理学 and 医学奖得主 ERIC R. KANDEL

发明PalmPilot的杰夫·霍金斯在本书中所展现出的科学素养和人文关怀精神，正是每一位载入史册的成功人士的秘诀所在。

——美国知名创业教练 JOHN NESHEIM

《人工智能的未来》辉煌而条理清晰，是几十年来在神经科学、心理学和人工智能领域的最重要的一部书。

——澳大利亚纽卡斯尔大学生物学教授 MALCOLM YOUNG

阅读此书，烧掉其他的书吧！杰夫·霍金斯将会改变世人对智能的看法和对智能机器的期待。

——风险投资公司KPCB的合伙人 JOHN DOERR

ISBN 7-5369-4018-1



9 787536 940185 >

责任编辑 龚毅华 林成岗

封面设计 阎谦君

ISBN 7-5369-4018-1/N · 39

定价：18.50元

人工智能 的 未来

杰夫·霍金斯

(Jeff Hawkins)

[美]

桑德拉·布拉克斯莉

(Sandra Blakeslee)

著

贺俊杰 李若子 杨倩 译

陕西科学技术出版社

图书在版编目(CIP)数据

人工智能的未来/(美)霍金斯(Hawkins, J.), (美)布拉克斯莉(Blakeslee, S.)著;贺俊杰,李若子,杨倩译. —西安:陕西科学技术出版社,2006.1

书名原文:On Intelligence

ISBN 7-5369-4018-1

I.人... II.①霍...②布...③贺...④李...⑤杨... III.人工智能
IV.TP18

中国版本图书馆 CIP 数据核字(2005)第 116759 号

ON INTELLIGENCE by Jeff Hawkins with Sandra Blakeslee

Copyright © 2004 by Jeff Hawkins and Sandra Blakeslee

Published by arrangement with author c/o Levine Greenberg Literary Agency, Inc.
through Bardon-Chinese Media Agency.

Simplified Chinese translation copyright © 2006 by Shaanxi Science & Technology Press
ALL RIGHTS RESERVED

陕西科学技术出版社享有本书中国大陆地区简体中文版专有权,该权利受法律保护,侵权必究。

人工智能的未来

出版者 陕西科学技术出版社

西安北大街 131 号 邮编 710003

电话(029)87211894 传真(029)87218236

<http://www.snstp.com>

发行者 陕西科学技术出版社

电话(029)87212206 87260001

印刷 西安信达雅印务有限公司

规格 880mm×1230mm 32 开本

印张 8.5

字数 185 千字

版次 2006 年 1 月第 1 版

2006 年 1 月第 1 次印刷

定价 18.50 元

(如有印装质量问题,请与我社发行部联系调换)

► 序 言 ◀

这本书的创作激情源自于我所钟情的两样东西——计算机和大脑。

25 年来我在移动计算机技术领域孜孜以求，毫无倦怠。在美国硅谷的高科技世界里，我可谓数得上的人物——曾创办过 Palm Computing 和 HandSpring 两家 PDA(个人数字助理或掌上电脑)公司；作为总设计师，也设计出许多掌上型计算机和智能电话，如 PalmPilot 和 Treo 等产品。

然而，我的第二种爱好不但早于我对计算机的兴趣，而且对它感情更为炽烈。我对大脑研究非常着迷。我想理解大脑是如何工作的，不仅仅从哲学的角度、从笼统的观点去理解，而是从细微处去探索，扎扎实实、精益求精地去理解。我不但渴望了解智能是什么、大脑又是如何工作的，而且更期望制造出像大脑那样工作的机器。创造出真正具有智慧的机器是我的最终目标。

当今，在地球上最后要攀登的科学高峰是解决智能的问题。大部分重大的科学论题所涉及的事件往往是极端微小、极端庞大，或者涉及亿万年前极为遥远的事件。然而，有关大脑的问题却不是这样。人人都有一颗大脑，你的大脑是属于你自己的。你为什么会这样感觉，而不是那样？你如何认知世界？为什么你会犯错？你怎样才能有创意？为什么音乐和艺术可以激发

灵感?如果想知道这些问题的答案,你必须了解大脑。另外,一个关于智能与大脑功能的成功理论不仅可以为治疗脑部疾病服务,还会带来巨大的社会效益。我们要建造真正的智能机器绝不同于科幻作品中所描述的机器人,相反,这种智能机器会随着一套有关智能本质的新理论大显身手。它们会帮助我们加速了解世界,辅佐人类探索宇宙,让世界变得更加太平。

我们有幸诞生在这样的时代,一个可以解答智能问题的时代。我们这一代人不但拥有百年以来收集到的堆积如山的关于大脑的数据,而且数据收集的速度也在日益加快。单在美国就有成千上万的神经科学家。可惜至今也没有形成一套有关智能和大脑工作原理的完整理论。大部分神经生物学家不大考虑有关大脑的理论,因为他们只贯注于所做的有关大脑的诸多子系统的实验,只埋头于收集更多的数据。尽管一批又一批的计算机程序师曾经尝试让计算机拥有智能,却一次次以失败而收场。我相信他们会屡战屡败,除非这些程序师不再忽视计算机与大脑之间的差别。

智能究竟是什么?为什么人脑有智能,而电脑没有?为什么一个6岁的小孩子可以在河床的石头上跳来跳去,姿态优美,而当今最先进的机器人却像行动迟缓的傻瓜?为什么3岁小孩已经粗通语言,而计算机却言语不通,枉费了半个世纪以来程序师艰苦卓绝的努力?为什么你能在1秒钟之内分辨猫与狗,而超级计算机却不能?这些都是亟待破解的斯芬克斯之谜。目前线索不少,但真正需要的是一些关键性的见解。

也许你很奇怪为什么一个计算机设计师要写一本关于大脑的书。换句话说,如果我对大脑的研究如此痴迷,为什么不将

脑科学或人工智能作为自己的事业呢?答案是,我尝试过,而且不止一次,但是我拒绝像前人那样研究智能问题。我相信解决问题的最好方法,就是利用人脑生物学提供的细节作为约束与指引,把智能当作一个计算性质的问题——这是一种介乎生物学与计算机科学之间的立场。许多生物学家基本上不接受或忽视以计算机术语的方式去认识大脑,而计算机科学家通常不相信生物学对他们来说有何借鉴之处。科学界比商界更不愿意冒险。在高科技行业,如果一个人从事新概念的研发工作,只要有充分的理由,无论成功与否,都能促进自己事业的发展。许多成功的企业家都是在早期失败之后才成功的。然而在学术界,如果新概念的研发工作经过几年时间仍没有成果,新生的事业就要夭折。所以,我决定同时追求生命中的两种爱好,并坚信商业上的成功会帮助我取得人脑研究的成功。我需要财力来支持我所追求的科学研究,我也需要学会推动世界变革的进程和推销新概念的方法,所有这些我在硅谷的工作中受益匪浅。

2002年8月,我创办了一家研究中心,取名为“红杉神经科学院”(Redwood Neuroscience Institute),简称RNI,专门进行脑理论的研究。世界上有很多神经科学中心,但唯此一家是专门从事全面的新大脑皮层理论研究,而新大脑皮层正是人脑中负责智能的部分。这就是我们在RNI的全部项目。

从很多方面来说,RNI就像一家初创公司。我们正在追逐有些人认为遥不可及的梦想,幸而这里人才济济,我们的努力已初结硕果。

这本书所表述的内容可谓雄心勃勃，它描述有关人脑如何工作的全面理论：智能是什么？人脑又是如何创造智能的？我提出的理论并不是全新的。你将读到的许多概念早已零散地存在，但将它们像珍珠般一颗颗串起来，构成一串光彩夺目的项链，本书却是首次。这也不足为奇。据说新概念往往是旧概念的重新包装和重新解释。这句话的确适用于这里提出的理论。然而包装和解释能产生天壤之别的效果，即仅有大量细节的素材与一个能够令人满意的理论之间的差别。我希望这个理论会打动你，使你产生类似的感觉。我听到的一种典型的反应是：

“有道理。我本来永远都不会这样看智能，现在既然你说了，我倒是能理解了。”一旦拥有这些知识，大部分人都开始用不同的眼光打量自己。你开始观察自己的行为，说：“我明白刚刚在我脑子里发生的一切。”希望读完这本书的时候，为何想其所想，又为何行其所行，你会对此有一个新的认识。也希望有些读者由此得到启发，根据书中勾画的原理，专注于建造智能机器的事业。

我常常把这个理论和我研究智能的思路称作“真智能”(real intelligence)，以示与“人工智能”的区别。人工智能科学家试图通过编写程序，让计算机模仿人类的行为，却没有回答智能是什么，其含义又是什么。他们遗漏了建造智能机器最重要的部分，即智能！“真智能”认为，试图建造智能机器之前，我们必须理解人脑是如何思考的，而这里并没有丝毫人工的东西。只有到此时，我们才可以考虑怎样建造智能机器。

本书前五章首先介绍为什么以前在理解智能与建造智能机器方面的努力都失败了，然后提出并进一步论述我称为记忆 - 预测架构 (memory-prediction framework) 理论的核心概念。第六章则细说人脑如何实现记忆 - 预测模型，换句话说，就是人脑究竟如何工作。接下来讨论这个理论对社会及其他方面的意义，对许多读者来说，这或许是书中最引人深思的章节。最后一章则讨论智能机器的建造与未来。但愿此书能让你神往。以下是我们将会一一探讨的问题：

计算机可以拥有智能吗？

数十年来，人工智能领域的科学家宣称，当计算机足够强大时，就可以拥有智能。我不这样认为，而且我会解释为什么。人脑和计算机的工作原理完全不同。

神经网络方面的研究会不会有助于智能机器的制造？

大脑固然由神经元组成的网络所构成，但若不首先理解大脑的工作原理，在创造智能机器方面仅凭简单的神经网络研究，将不会取得比开发计算机程序更大的成功。

破解人脑工作原理之谜为何如此困难？

多数科学家认为大脑是如此复杂而高深，我们需要花很长的时间才可以理解它。我不同意。复杂而高深是思想混乱的征兆，并不是其原因。反之，一些直觉的假设和错误的假设把我们误导了，其中最大的错误就是相信智能是以智能行为来定义的。

智能若不是以行为来定义，又该以什么来定义？

大脑使用大量的记忆来创造世界的模型。所有你知道的和学到的事物都储存在这个模型里。大脑用这个以记忆为基础的模型去不断地预测未来的事件。预测未来的能力才是智能的关键。我将深入地描述大脑的预测能力，它正是本书的核心概念。

人脑是怎样工作的？

智能产生于新大脑皮层。新大脑皮层尽管多才多艺而且适应力极强，然而它的结构细节却出奇地规则。新大脑皮层的不同部位，不管是负责视觉、听觉、触觉，还是语言的部分，都遵循相同的原理工作。理解新大脑皮层的关键在于理解这些共同的原理，特别是它的层级结构。我们将深入分析新大脑皮层，展示它的结构是怎样捕捉世界的。这些讨论将是全书中最具技术性的部分，但只要你对此感兴趣，即使你不是科学家也可以读懂。

这个智能理论的含义是什么？

有关大脑的理论可以帮助解释许多事物，例如我们怎样才有创造力，我们为什么会觉得自己有意识，我们为什么会持有偏见，我们是如何学习的，为什么有时却“老狗难学新把戏”，等等。我将会讨论许多这样的话题。总的来说，这个理论能帮助我们洞察自己是谁，又为何行其所行。

我们能够建造智能机器吗？它们又将会做什么？

是的，我们能，而且我们会。今后数十年，我预见这种机器的性能将会朝着有趣的方向迅速地演化。有人害怕智能机器会危害人类，但是我强烈地反对这种危言耸听的想法。人类是不会被机器人打垮的。建造在物理、数学等高层思维方面超越我们的机器，比建造科幻小说中会走、会说的机器人要容易得多。这种令人难以置信的技术正是我要探索的。

我的目标就是要以人人能懂的方式说明这个新的智能理论和人脑的工作原理。一个好的理论应该容易理解，不应该隐匿在专业术语或者纷繁复杂的论据之中。我将从一个基本架构出发，然后逐步添加细节，有的单纯是逻辑推论，有的则涉及人脑连接回路的特定方面。我所提议的某些细节肯定会有错，不过这在任何科学领域中都是难免的事。要发展一个完全成熟的理论需要多年，但这无损理论核心概念的力量。

多年前，当我最初对大脑发生兴趣的时候，我去当地的图书馆想找一本解释人脑如何工作的好书，那里有几乎所有有趣的主题，包括相对论、黑洞、魔术、数学等，这些都是当时令我神往的东西。十几岁时我早就习惯了找到有关的好书，然而却寻不到一本令人满意的说明大脑的好书。我终于认识到，没有人知道人脑实际上是如何工作的，甚至连一个基础的或者未经证实的理论都没有。这是不正常的。例如，当时虽然没有

知道恐龙是怎样灭绝的，但与之相关的理论却很多，而且全部都可以读到。关于人脑的情况则完全不一样。起先我不相信，我们竟然不知道这个关键的器官是怎样工作的。在对已知的有关大脑的研究过程中，我树立起一种信念：一定有个简明直接的解释。大脑不是魔术，而且我觉得答案甚至不会像魔术那么复杂。数学家保罗·厄多斯(Paul Erdos)相信，最简单的数学证明早已经写在“天书”里，数学家的工作就是解读天书，找出这些证明。同样地，我觉得智能的解释就在“天书”里。我能体验个中滋味。我想读懂那部“天书”。

25年来，我一直都有一个梦想，希望写出一本解释大脑的简明系统的小册子。它好像吊在我面前的一根胡萝卜，不断激励我。这个梦想成就了正握在你手中的这本书。我向来不喜欢复杂，无论是科学还是技术。这种偏好反映在我所设计的产品中，它们往往以容易使用而知名。最有力量的东西往往都是最简单的。这本书提出一个简明直接的智能理论。我想你会喜欢上这本书。



► 目 录 ◀

序 言

第一章 人工智能 (1)

有些科学家认为计算机可以做到人脑所能做的一切，人工智能机器的制造只是编写程序的问题。于是，一代又一代科学家们不断尝试，从 Eliza 程序到“积木世界”，再到著名的“中文屋”实验。我们清楚地看到，尽管这些程序可以模拟大脑，但它们不可能具有智能。直觉告诉我，这样的传统方式不可能制造出真正的智能机器。只有研究大脑、认识智能才是最终的解决之道。

第二章 神经网络 (17)

20 世纪 60 年代后期，一种新的智能机器研究途径崭露头角，它就是神经网络。相对于人工智能而言，这是一个进步，因为它是以真正的神经系统为基础的。但令人遗憾的是，神经网络的发展很快就固定于一些简单的模型之上而停滞不前。它对于时间和反馈的忽略使我进一步坚信，要制造出智能机器，必须从天然的智能引擎——新大脑皮层开始探索。

第三章 人脑 (35)

从各个角度观察人脑，其外表几乎没有什么区别，但人脑的内部包含许多不同的功能区，这些功能区在一个复杂的层级结构中相互联系。此外，神经科学家蒙卡斯尔认为，大脑皮层不仅在外表和结构上极其相似，进入大脑皮层的输入信息也是基本相同的，不论是听觉、视觉还是触觉。那么，我们的大脑到底是如何感知这个世界的呢？

第四章 记忆(61)

大脑皮层和计算机完全不同,它不会计算问题的答案,而是使用存储的记忆来解决问题,做出反应。由此,我们又接触到一些崭新的重要概念,如存储模式序列、自-联想记忆、恒定表征等等,它们发生在大脑皮层的每一个区域,无处不在。

第五章 智能理论新架构(83)

我们的大脑是用一种平行的方式连续不断地对我们所在环境的构架进行预测。这是大脑皮层的主要功能,也是智能的基础和必要因素。那么,我们的大脑皮层是如何进行预测的?

第六章 大脑皮层工作原理(105)

大脑皮层为了预测,必须能够记住并且存储关于事件序列的知识。要预测新的事件,它又必须形成恒定表征,而恒定表征存在于包括视觉、听觉、触觉和联合区等大脑的所有区域。大脑皮层区域到底是什么样子?皮层区域的详细工作原理是怎样的?为什么需要反馈呢?有了反馈就能解决问题了吗?大脑皮层是如何学习的?

除了记忆、反馈和恒定表征以外,大脑还需要按照世界的本来面貌建立并存储一个世界的模型。大脑存储世界的方式就是形成新的分类和新的序列。

第七章 意识和创造力(181)

创造力来自哪里?意识是什么?想像力又是什么?动物是否有智力?人的智力与动物的智力有何区别?什么是现实?这些问题的答案尽在本章中。

第八章 智能之未来(211)

很难预言一项新技术的最终用途，但设想一下未来智能机器的样子还是有所裨益的。人们能研制出智能机器吗？如果可以，那它们将会是什么样子呢？它们将会被用来干什么呢？这项技术是不是很危险，会不会危害到人类并对个人自由构成威胁？智能机器最显而易见的应用范围将是什么？智能机器对我们生活的最终影响将会是什么呢？

结语(243)

附录 可检验的预言(247)



第一章

人工
智能

Artificial Intelligence



当我 1979 年 6 月从康奈尔大学毕业时，我对自己的生活没有任何长远的打算。获得电机工程学位后，我在位于俄勒冈州波特兰市新建的英特尔工业园找到了一份工程师的工作。当时，微型计算机工业方兴未艾，而英特尔是这一领域的核心。单片机是我们当时的主导产品（由于英特尔发明了微处理器，因此在那时，把整台计算机建立在单个电路板上的想法就已经成为了可能）。而我的工作就是分析并修复由其他工程师发现的单片机上的错误。那期间，我创办了一份业务通讯，并经常外出旅行，有机会见识各种客户。尽管很思念在辛辛那提工作的大学女友，但那时的我年轻，过得很快活。

几个月后，我遇到了一件改变我人生目标的大事。那是一本当年 9 月新出版的《科学美国人》(Scientific American) 杂志，是有关大脑研究的专刊。它重又唤起了我自小就开始的对大脑的兴趣。这本杂志令人着迷。里面的文章涉及大脑的组织、发展和化学特征，专事视觉、运动及其他此类组织的神经机制，以及精神错乱的生理基础等内容。我认为它是《科学美国人》里最棒的一期。一些与我交往过的神经学家也表示，这期杂志在帮助他们做出职业选择的时候起到了重要作用。这一点和我一样。

这一期的最后一篇文章题为《有关大脑的思考》，作者是弗朗西斯·克瑞克(Francis Crick)，他曾是 DNA 结构的发现人之一，并且他那时就已经将其天才思考投入到了对大脑的研究之中。克瑞克认为，尽管一直以来积累了大量有关大脑的详尽知识，但大脑的工作原理对于人们来说仍是一个难解的谜。一般来说，科学家不会写他们不知道的事情，但克瑞克就像那个指出皇帝没有穿衣服的小男孩一样，对此并不在意。他认为神

经学只是一堆没有任何理论的数据。他说：“最明显的是在概念上缺乏一个总的框架。”这句话在我听来就像一位英国绅士在委婉地表示：“我们对大脑的工作原理一无所知。”而那时这是事实，并且现在也是如此。

克瑞克的话像是出发的号角声，唤醒了我长久以来梦想研究大脑、制造智能机器的愿望。尽管我已走出大学校园，但还是决定做出事业上的转变——我打算研究大脑，不仅仅是理解大脑是如何工作的，而且要以这些知识作为新技术的基础，制造出智能机器。但要把这一计划付诸实践尚需时日。

1980年春天，我调到了英特尔驻波士顿的公司，和我未来的妻子团聚了，她那时正在当地的学校攻读硕士学位。我的工作是客户和员工们讲解如何设计微处理器系统，但我心中的抱负却在另一方面——我一直在寻找研究大脑理论的方法。所学的工程学知识使我意识到，一旦解开了大脑工作原理之谜，我们就可以复制整个过程，而且硅是制造人工大脑的天然材料。因为我就职的公司发明了硅记忆芯片和微处理器，因此我想，也许能说服英特尔，允许我把一部分时间用在智力研究和设计仿大脑的记忆芯片上。于是我给英特尔的主席戈登·摩尔(Gordon Moore)先生写了一封信，大意如下：

亲爱的摩尔先生：

兹建议成立一研究小组，专攻大脑工作原理的研究。该工作可从一个人，即本人开始，随后进一步拓展。本人有信心承担该工作。相信有一天它会给我们带来无限商机。

杰夫·霍金斯

随后，摩尔把我介绍给英特尔的首席科学家特德·霍夫(Ted Hoff)。我飞到加利福尼亚与他会面并陈述了对大脑研究的建议。霍夫在两个方面成绩卓著：一个是他在第一台微处理器的设计工作中表现出色；另一个是进行了神经系统理论的早期研究。前者为我所熟知，而后者在当时我还不甚了了。霍夫曾做过人造神经元及其实际应用等方面的研究，而这些都是我所未曾想到的。听完我的建议后，他表示不相信人类能够在可预见的未来研究出大脑的工作原理，因此英特尔公司没有理由支持我的想法。他的看法是对的，因为在25年后的今天我才刚刚开始在大脑的研究方面取得显著的进展。我很清楚，在商界时机就是金钱，不过当时我仍然感到很失望。

原本我想通过一条便捷的道路达到目的，而英特尔公司应该是进行大脑研究最便利的桥梁。但此路不通，我只得另辟蹊径。于是我考虑申请麻省理工学院的研究生院，因为该学院以人工智能研究著称，而这将对我的研究提供很大的便利。幸运的是，我似乎完全具备申请的条件：接受过计算机科学的专门训练——“合格”；立志建造智能机器——“合格”；以大脑研究作为突破口来认识其工作原理……“且慢，这是个问题。”这最后一项——弄清大脑工作的原理——在麻省理工学院人工智能实验室的科学家们看来是个问题。

而难题接踵而至。麻省理工学院是人工智能的航母，在我申请的同时，已有大批精英汇聚于此，意欲通过程序设计使电脑具有智能。在这些科学家看来，视觉、语言、机器人科学和数学都只是编写程序的问题。既然计算机可以做到人脑所做的一切，甚至更多，那为什么我们的思维还要大脑——这一天然

计算机的生理混乱所限制呢？他们认为研究大脑会限制思维，更明智的做法应该是发掘计算方法的最大极限。而这一点已经被数字计算机发挥得淋漓尽致。他们所追求的就是编写出能与人脑相媲美并最终超越人脑的程序。他们对大脑如何工作毫无兴趣，采取“只求结果，不问手段”的方式开展研究。甚至有些人还为自己跳开了神经生物学这一阶段而沾沾自喜。

确切地说，我觉得这是一种错误的解决问题的办法。直觉上我感到这种人工智能的方法不仅不能设计出能与人脑相媲美的程序，甚至不能告诉我们智能到底是什么。计算机和大脑的工作原理完全不同：一个是设计出来的，一个具有自学能力；一个必须不断完善才能发挥作用，一个天生可以灵活地处理和宽容失误；一个有中央处理器，一个则没有中央控制……这样的差别举不胜举。我之所以认为计算机无法智能化，主要原因是，从晶体物理学的层面来看，我明白电脑是如何工作的，而恰恰是这一点使我意识到大脑与计算机在本质上完全不同。对此我无法证明，但直觉告诉我这是对的，就像人们对一些事情有直觉一样。最终我得出结论：用传统的方式研究出的人工智能可以生产出实用的产品，但绝不可能制造出真正的智能机器。

与此相对，我想廓清什么是真正的智能和知觉，探究大脑生理学和解剖学理论，挑战弗朗西斯·克瑞克提出的论断，为大脑的工作原理绘制出一个大的框架。我把精力集中在对新大脑皮层的研究上——它是近期哺乳动物大脑进化最快的一个部分，也是智能的中心。只有认识了新大脑皮层的工作原理之后，我们才能着手建造智能机器，而在此之前是不可能做到这一点的。

令人遗憾的是，我在麻省理工学院遇到的教授和学生无人对此感兴趣。他们直截了当地告诉我，要认识智力和建造智能机器，没有必要研究真正的大脑。1981 年该大学拒绝了我的申请。

今天，许多人认为人工智能是一个既成的事物，只待计算机具有足够的能力后，即可实现其美好前景。只要计算机有足够的内存和处理能力，思维能力就会发展，人工智能的编程人员就可以造出智能机器。对此我不敢苟同。人工智能正面临着一个根本的错误，因为它无法圆满地解决什么是智能的问题，或者说“理解某个事物”到底意味着什么。回顾人工智能的发展史及其建立的原则，我们可以看到这一领域的发展偏离了正确的方向。

人工智能理论是随着数字计算机的出现应运而生的。早期人工智能研究的关键人物是英国数学家阿兰·图灵(Alan Turing)——多用途计算机这一想法的发明人之一。他最卓越的成就在于正式提出了“万能计算”这一概念，即：尽管建构的细节有所不同，但从根本上讲所有的计算机都是等效的。他还假想了一台机器来证明这一观点。这台虚构的机器包括三个主要部分：一个处理盒，一个纸带和一个能从来回移动的纸带上读取并记录信息的装置。纸带就像众所周知的计算机代码 0 和 1 一样，是用来存储信息的(当时没有发明记忆芯片和光盘驱动器，所以图灵想像用纸带来存储信息)。那个处理盒，现在被称为中央处理器(CPU)，会遵循一定的规则读取和编辑纸带上的信息。图灵用数学的方法证明，如果 CPU 的规则选择正确，并

且有足够长的纸带供其使用，这台机器就可以完成世界上任何一种计算。这就是被称为“图灵万能机器”的众多等效计算机中的一种。无论是计算平方根，还是弹道轨迹，是玩游戏、编辑照片，还是做银行交易核对，其基本的形式都是0和1，任何一台“图灵机”都可以处理。信息处理就是信息处理。从逻辑上讲，所有的数字计算机都是等效的。

图灵的结论有着无可辩驳的正确性，取得了惊人的成功，计算机及其所有产品都以之为基础。随后，图灵转而考虑如何建造智能机器。他感到电脑可以智能化，但他不愿卷入对其可能性的争论之中，并且认为自己无法给“智能”下一个定义。因此他在这一方面没有做任何尝试，而是提出了一个智能存在的证据，即著名的图灵实验：如果一台电脑可以诱使一个问询者相信它也是一个人，那么从定义上来讲这台机器就有智能。以这个实验为衡量尺度并以“图灵机”作为媒介，图灵开创了人工智能的研究领域，其核心是：大脑就是另一种形式的计算机，无论你怎么设计它的人工智能系统，这台计算机都可以做出类似人类的行为。

人工智能的倡导者们看到了电脑与思维之间的相似之处。他们说：瞧，很显然，人类智力最伟大的功绩就在于对抽象符号的处理——而计算机也能做到。所谓听和说的过程，就是我们用恰当的语法规则对被称为“词”的概念符号的处理过程。下棋时会怎样？是在处理那些代表着不同棋子的属性和位置的概念符号。那么看东西呢？我们用概念符号代表物体以及物体的位置、名称及其他属性。当然，人们是借助大脑而不是计算机来做到这一切的，但图灵已经证明，如何实施和处理这些符号并

不重要，你可以借助一个齿轮装置、一个电子系统或是大脑神经元网络等等，任何形式都可以做到，只要你的媒介可以实现和图灵万能机器相同的功能即可。

这一假想得到了一篇有影响的科学论文的支持。该论文发表于 1943 年，作者是神经生理学家沃伦·麦卡洛克 (Warren McCulloch) 和数学家沃尔特·皮兹 (Walter Pitts)。他们俩解释了神经元是如何数字化工作的——也就是说，大脑的神经元是如何能够对规范逻辑进行充满想像力地复制。他们认为，大脑神经元的工作原理和电脑工程师所说的逻辑门一样。而逻辑门可以处理简单的逻辑关系，如“并且”“否”和“或者”。计算机的芯片是由上百万个逻辑门构成，它们被合成精确而复杂的电路。一个 CPU 就是一个逻辑门的集合体。

麦卡洛克和皮兹指出，用精确的方式可以将神经元连接起来，以实现逻辑功能。神经元可以收集并处理输入的信息，从而决定是否输出信息，因此可以将神经元想像成一种活的逻辑门。这样，我们可以大胆地设想，大脑是由“并且”“或者”等逻辑门及其他此类的逻辑节点构成的，和数字电路相类似。我们不清楚麦卡洛克和皮兹是否真的相信大脑是这样工作的，他们只是说这大有可能。而且从逻辑上讲这种看法是可能的，因为神经元在理论上可以实施数字功能。而后，再没有人专门针对神经元在大脑中的存在形式提出疑问。尽管缺乏生物学的根据，他们想以此作为证据来证明，大脑是另一种形式的计算机。

还有一点也是值得一提的：人工智能的基本原理得到了一个在 20 世纪上半叶占主导地位的心理思潮的支持，人们称之为“行为主义”。行为学家认为，大脑内部的运转过程是不可

知的，称之为“无法穿越的黑匣子”；但是，人们可以观察和测量动物所处的环境和行为，例如它们如何感知，行为如何等。他们承认大脑具有反射机制，可以利用奖惩的方式使动物学会新的行为。但除此之外，人们没有必要研究大脑，尤其是那些杂乱无章的主观情感，如饥饿、恐惧或什么是“理解”，等等。毋庸置疑的是，这种实验哲学在 20 世纪后半叶逐渐消亡了，而科学家对人工智能的兴趣却没有因此而消亡。

二战结束后，电子数字计算机的广泛应用已成为可能，于是人工智能的先锋们开始摩拳擦掌设计程序。语言和语言之间的翻译？简单！只是破译代码而已，只需将 A 系统中的每一个代码和 B 系统中的代码相匹配即可。图像呢？也很简单。已经了解了处理图像旋转、增减和移位所需的几何定理，只要将它们编成计算机代码即可。这样，工程就已完成了一半。人工智能专家曾骄傲地断言，计算机的智力将很快赶上并超过人类。

具有讽刺意味的是，有一种名为“Eliza”的电脑程序几乎可以达到图灵实验的要求，它能够改变你的措辞从而模仿精神分析师。例如，如果一个人输入“我和我的男朋友不再说话了”，Eliza 就会说：“跟我说说你的男朋友。”或者“你和男朋友为什么不说话了？”这一程序设计平庸且不全面，初衷只是一个玩笑，但还是蒙骗了一些人。除此之外，还出现了一些严肃的程序，如“积木世界”。该程序模拟了一个房间，里面装有各种颜色和形状的积木。你可以提出各种问题，如“绿色金字塔形的积木是在红色大方块儿的上面吗？”或“请把蓝色的方块移到红色小方块儿上。”该程序就可以回答你的问题或是按照你的要求完成动作。这一切都是模拟的，虽然有效，但只能

局限在这个完全人造的小小积木世界里。程序员不可能将其扩展到任何实际的用途中去。

与此同时，公众被这一连串表面上的成功以及有关人工智能技术的新闻消息深深吸引。最初只是引起人们兴奋的程序竟然可以解决数学定理的问题了。自柏拉图以来，多步骤演绎推理一直被认为是人类智慧的顶峰，因此，在最初看来，人工智能似乎有很大的经济潜力。但结果证明，就像“积木世界”一样，这个程序只是能够找到简单定理而已，而且这些定理早已为人所熟知。随后一个名为“专家系统”的程序引起了人们的巨大关注。它是一个包含细节事实的数据库，能够回答人提出的问题。比如“医学专家系统”就能根据列出的症状诊断出病症。但遗憾的是，这个系统再一次被证明作用有限，无法显示出与综合智力有关的任何东西。除此之外，计算机还可以具有大师级的象棋技艺，而且 IBM 的国际象棋机器人“深蓝”战胜了国际象棋世界冠军加里·卡斯帕罗夫 (Gary Kasparov)，曾经轰动一时。但这种胜利是毫无意义的，因为“深蓝”的取胜不是因为它比人聪明，只是在于它的速度是人脑速度的几百万倍。“深蓝”没有直觉，而一位象棋大师综观棋盘上的局势，立即就可以判断出比赛中哪个区域有利，哪个区域有危险。然而电脑对于重要的事情没有任何直觉的感应，它必须尝试更多的选择。此外，“深蓝”对于这种游戏的历史毫无了解，对自己的对手也一无所知。它就像一台计算器，能算算术但毫不理解什么是数学。“深蓝”只能下棋，并不真正了解象棋。

在任何情况下，无论多么成功的人工智能程序也只是擅长于那些经过专门设计的领域。它们不会总结归纳，缺乏灵活

性；甚至它们的创造者也认为它们不会像人一样思考。当初被认为很简单的人工智能问题最终都无果而终，至今也没有一台计算机能够在语言理解方面超过 3 岁的儿童或是在视觉方面超过任何一只老鼠。

多年的努力带来的只是无法兑现的承诺和毫无说服力的成果。人工智能头上的光环开始慢慢退去，一些科学家转而进入其他领域的研究。刚刚成立的人工智能公司纷纷倒闭，资金来源越来越少。一时间，使计算机具备最基本的感知、语言和行为能力的程序设计似乎都变得不再可能。直到今天，这一境况都没有大的改变。正如我先前所说，尽管有人相信可以通过速度更快的计算机来解决人工智能的问题，但大多数科学家都认为以前所有的努力都是有缺陷的。

我们不应因此而责怪人工智能研究的先驱们，毕竟，阿兰·图灵是伟大的。这一切能够说明，图灵机将会改变世界——而且它已经改变了世界，但不是通过人工智能。

我申请麻省理工学院时就对人工智能的断言心怀疑虑。当时，加州大学伯克利分校的知名哲学教授约翰·塞尔(John Searle)就曾说，计算机没有，而且也不可能有智能。为了证明这一观点，他于 1980 年设计了一个思考实验，即“中文屋”实验。

假设一个说英语的人坐在一间墙上开有小孔的房间里，在他面前的桌上有一本指令手册和各种可能需要的铅笔和便签纸，翻开手册，他可以看到用英语写成的指令，其中介绍了汉字处理、分类及比较的方法。请注意：这些指令没有提及任何

汉字的含义，只是描述如何对汉字进行复制、删除、重新排列和誊写等内容。

房间外的一个人从小孔中递进一张纸，上面用中文写着一个故事和与之有关的问题。屋内的人对中文一窍不通，但他在手册的指令下拿起笔开始工作——只是生搬硬套地按照指令去做。有时指令要求他在纸上写下一些汉字，有时又让他移动或删除另外的汉字。就这样写写画画直到指令告诉他工作已经完成。这时他已写出了一篇汉语文章，而他并不知道，这正是那些问题的答案。他按照指令将纸从小孔送出去，心中疑惑，这无聊的游戏到底是什么。

房间外一个懂汉语的人读过文章后说答案完全正确，而且很有见地。如果问她(懂汉语的人)，这篇文章是否出自一个对故事透彻理解的聪明人之手，她一定会很肯定地回答“是”。她对吗？是谁看懂了故事？当然不是屋里的那个人——他根本不懂中文，因而对故事一无所知；也不是指令手册——它只是躺在桌上一堆纸张中间毫无生命的书。那么，理解从何而来？塞尔的回答是：根本不存在所谓的“理解”——存在的只是无意识的书页翻动和铅笔的涂鸦而已。这个“中文屋”与数字计算机非常相似，那个人就是 CPU——只是无意识地执行指令；那本书就是向 CPU 传达指令的软件程序；而那些便签纸便是内存。因此，无论你多么巧妙地进行设计，试图让计算机通过模仿人类的行为而积累一定的智力，计算机都是不可能具有理解力和智能的。(塞尔曾明确表示他不清楚智能是什么，他的意思是不论智能是什么，电脑都不具备。)

这个论点引发了哲学家和人工智能专家之间的激烈争论。

他们发表了上百篇文章，字里行间夹杂着尖刻和愤怒。人工智能的捍卫者们提出了一些驳斥塞尔的论据，如：虽然房间中的任何部分都不懂中文，但房间作为一个整体是理解中文的；房间里的人是懂中文的，但他自己对此没有意识，等等。就我而言，我认为塞尔的解释是正确的。认真思考过“中文屋”实验的论据和计算机的工作原理之后，我没有看到任何地方有“理解”的发生，这使我坚信，我们首先需要弄清楚什么是“理解”，并为之下一个定义，这样才能清楚地知道何时一个系统是智能的，何时不是；何时它能懂中文，何时不能。单从它表面的行为是不能找到答案的。

一个人要看懂一个故事是不需要刻意去做任何事的。我可以静静地读着一个故事，表面上没有任何行为表明我是否清楚地理解了，至少对我而言是这样。而另一方面，你无法从我安静的行为上辨别出我是否读懂了故事，甚至无法辨别写成故事的语言我是否能懂。事后你可以提问，但我对故事的理解发生在阅读之时，而不是你问话之际。这本书的主题就是：理解是无法用外部行为测量的，它是对大脑如何记忆、如何利用这些记忆进行判断的内部度量——这些将在以后的章节中写到。在这一点上，“中文屋”“深蓝”和大部分的电脑程序一样没有任何类似之处，它们不能理解正在做的事情。而电脑的输出，即外部行为，成为我们判断它是否具有智能的唯一标准。

人工智能派为自己辩护的终极论据是：从理论上讲，计算机可以模拟整个大脑。一旦计算机模拟了所有的神经元和它们之间的连接点，就说明大脑的“智力”和计算机的模拟“智力”不再有任何区别。尽管这在实际上不能成立，但我同意这

种看法。可惜的是，人工智能的研究者们并没有模拟大脑，因而他们设计的程序不会具有智能。在不了解大脑是什么的前提下模拟大脑，是不可能的。

被英特尔和麻省理工学院拒绝后，我无所适从。如果你不知道下一步该干什么，最好的策略就是保持现状直到有转机出现。于是我继续从事电脑行业，心满意足地待在波士顿。直到1982年，我妻子想移居加州，于是我们举家前往(这是避免摩擦的一个办法)。我在硅谷找到了工作，是一家刚刚成立的小公司“Grid Systems”。Grid是便携式电脑的发明人。而便携式电脑是一台绝妙的机器，后来成为纽约现代艺术馆中第一个电脑藏品。我先后在市场部和工程技术部工作，后来发明了一种高级程序语言，命名为“Grid Task”。我和我的发明对于公司的成功变得越来越重要，同时我的事业蒸蒸日上。

然而此时，对于大脑和智能机器的好奇仍然在我脑海中挥之不去，我一心渴望研究大脑，于是参加了人体生理学的函授课程，开始自学(没有任何人会被函授学校拒绝的)。学习了一定的生物学知识后，我决定申请生物学研究项目的研究生，希望从生物科学的内部去研究智能。如果说计算机世界不需要大脑研究的理论家，那么生物学界可能会欢迎一位计算机工作者。那时还没有所谓的理论生物学，尤其是理论神经科学，所以从我的兴趣来看，生物物理学是最好的选择。我努力学习，参加了入学考试，准备了简历，恳请公司写了推荐信……哇，最终我被接收为加州大学伯克利分校的生物物理学研究项

目的全日制研究生。

我欣喜若狂。终于可以认真地研究大脑理论了，我这样想。我辞了 Grid 公司的工作，决心不再重回计算机产业任职。当然，这意味着无限期地放弃我的薪水。那时我妻子正盘算着买房子和生孩子，而我却心甘情愿地变成了一个不能挣钱养家的人。这可不是避免摩擦的好办法，但这是我最好的选择，而且我的妻子也支持我的决定。

在我离开公司之前，Grid 的创始人约翰·艾伦比(John Allenby)把我拉进办公室，说：“我知道你不想再回 Grid，也不想再涉足电脑业，但以后会发生什么谁也不知道。不要完全退出，就算休假吧。这样，一两年后如果你回来，还可以重领薪水，继续做你的职位，得到公司的股份。”这是一个友好的表示，我接受了，但心中仍然感觉这次我永远告别计算机产业了。



第二章

神经网络

Neural Networks



1986年1月，我开始了在加州大学伯克利分校的学习。期间，我所做的第一件事就是整理、汇编有关智能和大脑功能理论的研究成果。我阅读了上百篇论文，它们的作者有解剖学家、生理学家、哲学家、语言学家、计算机科学家和心理学家。许多来自于不同领域的研究者发表了他们对思维和智能的看法，各个领域都有自己的相关刊物和术语，尽管如此，我还是发现他们对思维和智能的解释有前后矛盾和不圆满之处。语言学家谈到智力时总会用“句法”“语义学”等术语，在他们看来，大脑和智力都和语言有关。视觉科学家总是会谈到2D、 $2\frac{1}{2}$ D和3D图画，对他们来说，大脑和智力与直观模式识别有关。计算机科学家总是谈及图式和框架等显示学识的新名词，他们中没有人提到大脑的构造，也没有人提到大脑是如何来实施这些理论的。另一方面，解剖学家和神经生理学家撰写了大量关于大脑构造和神经作用机理的文章，但对于大规模理论的提出，他们都采取了回避态度，因为要把各种研究方法以及与其相配套的堆积如山的实验数据弄清楚，实在是一件令人头疼的事。

正在此时，另一种智能机器的研究途径浮出了水面，这一次是全新的，而且充满了希望。早在20世纪60年代后期，神经网络就以某种面目出现了，但它和人工智能的研究在投资强度和投资方的关注度方面是死对头。那时，人工智能就像一只重达350千克的大猩猩，可以轻而易举地将神经网络研究碾得粉碎。神经网络的研究者们在此后的几年中一直处在投资方的黑名单上，只有少数人坚持了下来，直到80年代中期，才得以重见天日。我们无法确切地知道神经网络何以会突然引起人们的兴趣，但有一点是毋庸置疑的：人工智能的不断失利是一个

主要因素。人们在极力寻找人工智能的替代品，最终在神经网络领域得偿所愿。

相对于人工智能而言，神经网络是一个真正的进步，因为它的基础是建立在真正的神经系统之上的，尽管不甚牢固。它的研究人员，又被称作连接主义者。他们摒弃了编程计算机，将兴趣转向另一个方面。他们希望探究的是：如果将一簇神经元连接在一起会产生什么样的行为。大脑是由神经元组成的，因此大脑就是一个神经网络，这是一个不争的事实。而连接主义者希望通过对神经元相互作用的研究，让智能不可捉摸的特性变得清晰，并且希望通过复制众多神经元之间的连接，让一些令人工智能束手无策的问题得以解决。神经网络不同于计算机，它没有 CPU，不能对信息进行中央存储；整个网络的知识 and 记忆都分散在所有的连接之上——就像真正的大脑一样。

表面看来，神经网络似乎迎合了我的兴趣，但很快我就对这一领域失望了。那时我头脑中已经形成了一个想法，即：对于大脑的解读，有三种标准必不可少。

第一个标准是大脑功能的时间概念。真正的大脑快速地处理大量不断变化的信息流，在这些输入和输出的信息流中没有任何东西是静止不动的。

第二个标准是反馈的重要性。神经解剖学家早已发现大脑是浸润在反馈连接之中的，比如说，在连接新大脑皮层和视丘（一个位置较低的结构）之间的神经中，反向传递的信息（与输入相对）要比向前传递的信息大一个数量级。也就是说，如果有 1 根神经纤维向大脑皮层输入信息，就有 10 根神经纤维在向感觉器官传递信息。反馈也通过大脑皮层控制了大部分的连接。虽

然反馈的确切作用还不为人所知，但从已发表的一些研究可以看出它们无处不在。因此我感到反馈是至关重要的。

第三个标准就是任何大脑模型或理论都应该能够解释大脑的物理结构。新大脑皮层并不是一个简单的构造，正如大家将要读到的，它是由一个不断重复的层级结构组成，任何神经网络如果对这一结构认识不足的话，就无法像大脑一样工作。

然而，让人感到遗憾的是，神经网络现象一经登台亮相，就固定在一些极为简单的模型之上，而这些模型与上面提到的三个标准不相吻合。绝大多数神经网络都包含少量相互连接的神经元，它们排成三列。当一种模式(输入)显示在第一列上，这些输入神经元就会与下一列神经元相连接，形成所谓的“隐藏单元”；接下来，隐藏单元就会与最后一列神经元连接，即输出神经元。这些神经元之间的连接强度不同，也就是说，一个神经元上的活动可能会促进第二个神经元的活动或者削弱第三个神经元的活动，具体情况视其连接强度而定。通过这些强度的改变，网络就能学会将输入方式映射到输出方式上。

这些简单的神经网络只能处理静止不动的模式，没有涉及反馈，和大脑也没有任何相似之处。有一种最普通的神经网络，被称为“反向传播(back propagation)”网络，可以学会将输出单元的错误传输回输入单元，并将它记住。你可能认为这是一种反馈，而事实并非如此。这种错误的反向传播只会发生在学习过程中，而当经过训练的神经网络工作正常时，信息只会向一个方向传输，从输出到输入没有产生反馈。除此之外，这些模型没有时间概念。当一种静止的输入形式转化成静止的输出形式，另一种输入形式就会显现出来。这个过程，哪怕是刚刚发生的，在网络

中都不会留下任何记录。最后,与大脑的复杂及其层级结构相比,这种神经网络的构造就显得太微不足道了。

我本以为这个领域会迅速地向更仿真的网络发展,但它没有。因为这些简单的神经网络可以完成一些有趣的任务,因而在以后的几年中一直停滞不前。正是出现了这种全新的有趣工具,一夜之间,上千的科学家、工程师和学生凭借它获得了助学金、博士学位或发表了著作;一个个公司纷纷成立,利用神经网络进行股票市场预测、处理贷款申请、核对笔记以及处理上百种各种其他形式的分类申请。虽然创建者的初衷可能是更为广阔的领域,但那时这一领域已经被另一些人控制,他们对大脑的工作原理以及什么是智能毫无兴趣。

大众媒体对这个差异也不甚了了。各种报纸、杂志和电视科学节目都将神经网络描述成“像大脑一样”或“是以大脑原理”工作的。与处处需要编程的人工智能不同,神经网络是通过事例学习的,这一点似乎更具智能化。其中一个主要代表就是 NetTalk,它能学会将字母序列和读音相匹配。一开始这个网络是用印刷文本训练的,因此听起来就是电脑的声音在朗读单词。不难想像,用不了多长时间,神经网络就可以和人类对话了。而全国新闻在报道时,却将 NetTalk 错误地解释成一种可以学习阅读的机器。尽管它的展示表演很精彩,但其实际功效只能归于微不足道——它不会读,不能理解,没有任何实用价值。它所做的只是将字母和预先确定的声音模式相结合而已。

让我们分析一下,看看神经网络与真正的大脑之间的差距到底有多大。假设我们要研究的不是大脑,而是一台数字计算机的工作原理。多年努力之后,我们发现电脑是由晶体管构成,上千

万的晶体管精确而复杂地连接在一起构成一台计算机，但我们还是不明白计算机是如何工作的，也不知道这些晶体管为什么要这样连接。于是我们决定将几只晶体管连在一起，看看到底会发生什么。你瞧，区区三只晶体管用某种方式连接在一起，就成了放大器，一端输入的信号在另一端被放大。（收音机和电视机里的放大器就是这样制成的。）这是一个伟大的发现，一夜之间形成了一项新工业，开始制造收音机、电视机和其他各种使用放大器的电器。这本无可厚非，但它并没能告诉我们计算机是如何工作的。尽管放大器和计算机都是由晶体管构成，但除此之外，它们再无任何相似之处。同样，真正的大脑和三列的神经网络都是由神经元构成，但它们几乎完全不同。

我在 1987 年夏天的经历，又给我对神经网络并不高涨的兴趣泼了一盆凉水。在一次神经网络会议上，一家名为 Nestor 的公司展示了一种能够识别模板手写文字的神经网络，并要价 100 万美元。这引起了我的注意。尽管他们在这种神经网络规则系统的尖端性上大做文章，甚至将它吹嘘成另一项重大突破，但我觉得手写识别的问题可以通过另一种更简单、更传统的方法解决。回家后，这个问题一直萦绕在我的脑海之中。两天后，我设计出一种速度更快、体积更小、使用更灵活的手写识别器。它不使用神经网络，工作原理也和大脑完全不同，其规则系统的灵感来自于我正在学习的一种与大脑有关的数学。那次会议促使我去设计一种带有触笔接口的电脑（最终这个想法成就了 10 年后的 PalmPilot 掌上电脑）。我设计的手写识别器为后来一种名为 Graffiti 的输入系统奠定了基础，这一系统在第一代 Palm 产品上广泛使用。我知道 Nestor 公司在这一场竞争中被淘汰出局了。

简单的神经网络就到此为止。它的大部分功能都可以被其他方式轻易地取代，媒体的大肆宣传最终也偃旗息鼓了。至少，神经网络的研究者们并没有表示他们的模型是智能的，它们毕竟只是一些极其简单的网络，与人工智能相比并无过人之处。我并不想给大家留下一种印象，认为所有的神经网络都只有三层简单的变化。实际上，一些研究人员一直在研究不同设计的神经网络。如今，这个名词被用来描述各种不同形式的模型，其中一些在生物学角度看来是精确的，而另一些则不是。但不论怎样，它们几乎都忽视了对新大脑皮层的功能和构造的研究。

我认为，大多数神经网络和人工智能都有一个共同的特点——它们只注重行为，而这正是神经网络的问题所在。不论这些行为被称为“回应”“模式”还是“输出”，人工智能和神经网络研究者都认为智能存在于行为之中，而这种行为是执行一个输入后，由一个程序或神经网络产生出来的。电脑程序和神经网络最重要的属性就是能否进行正确的、令人满意的输出，就像阿兰·图灵所说的，智能等同于行为。

而实际上，智能并不是一种动作，也不是某种聪明的行为。行为只是智能的一种表现，绝不是智能的主要特征或对于智能的首要解释。思考就是有力的证明：当你躺在黑暗中思考时，你就是智能的。如果忽略了头脑中的活动而只关心行为，将对智能的理解和智能机器的发明造成障碍。

在进一步探索智能的定义之前，我想介绍另一种连接方式，它与真正大脑的工作原理更接近，但问题是，几乎没有人

认识到这项研究的重要性。

正当神经网络大出风头时，一小部分理论学家从中分离出来，发明了一种不以行为为中心的网络，称为“自-匹配记忆”（auto-associative memories）。它也是由相互连接的简单神经元构成，这些神经元在达到某个临界点时可以自动激活。但它们之间的连接方式不同，其中涉及大量的反馈。与只能正向传输的反向传播网络不同，自-匹配记忆可以将各个神经元的输出传回给输入——就像给自己打电话。这种反馈回路形成了一些有趣的特点。当一种行为模式施加在人造神经元上之后，它们就会对这种模式形成记忆，这种记忆网络将行为模式和它自身联系在一起，因此被称为“自-匹配记忆”。

乍看起来，这种回路的结果似乎很荒谬。如果要检索一个被存储的模式，你就必须先提供这个模式。这就像去食杂店买香蕉，当店主问你如何付款时，你回答说用香蕉付账。你可能会问：“这有什么用呢？”然而，自-匹配记忆所具有的一些特征恰巧能在大脑中找到。

第一点，其最重要的特征是：要检索某个模式，不必要有这个模式的全部，只要有其中的一个部分或一个乱作一团的模式，即使你从一个混乱的形式开始，自-匹配记忆都可以检索到以原始形式存储的正确模式。就像你拿着一些吃剩的、已不新鲜的香蕉来到杂食店，换回的却是整爪的新鲜香蕉。或是你拿着被撕破、已无法辨认的钞票来到银行，工作人员对你说：“我想这是一张破损的 100 元钞票。把它给我，我给你换一张崭新的 100 元。”

第二点，和其他大部分神经网络不同，自-匹配记忆可以

存储多序列模式，也被称作时间模式。要实现这一特征只要在反馈上设计一个延时即可，这样你就可以输入一个序列模式，比如一段旋律，自-匹配记忆就可以记住它。当我输入“一闪一闪小星星”的前几个音符时，它马上就可以播出整首曲子。如果输入的是一个序列的一部分，它也能将其他部分回忆起来。正如我们将要读到的，在现实中，人就是用这样的序列模式进行学习的。我认为大脑就是利用了一个类似自-匹配记忆的回路来实现这一目标的。

自-匹配记忆反映了反馈和随时间变化的输入的重要性。但遗憾的是，绝大多数的人工智能、神经网络以及认知科学家都没有注意到时间和反馈。

从整体来看，神经科学家的研究也不尽如人意。尽管他们知道反馈，而且反馈也是他们发现的，但他们当中大多数人没能够提出任何理论(除了对阶段和调整有不确切的看法之外)对大脑为何需要如此多的反馈做出解释，而且在他们对于大脑完全功能的看法中，时间所占的地位也是微乎其微。他们说明大脑活动的依据是某个事件在大脑中所处的位置，而不是神经激活模式相互作用的时间和方式。之所以出现这种偏见，部分原因是目前实验技术的限制。20世纪90年代——“大脑研究的十年”中，最令人振奋的技术之一是功能造影。功能造影仪可以给人脑活动拍下照片，但却无法记录快速的变化。于是科学家们就像拍普通光学照片那样，要求受试者一遍又一遍地将注意力集中在单一任务上。但这不是普通光学照片，而是智力活动的照片。结果，他们获得了大量数据，表明某种任务在大脑的反映区域，但有关随时间变化的输入是如何流经大脑的数据

却几乎没有。功能造影技术使我们观察到特定时刻事物发生的区域，却无法解释大脑活动在一段时间内是如何变化的。科学家希望收集到这样的数据，但苦于没有可行的技术支持。于是，许多主流的认知神经科学家仍然相信那种输入/输出谬误，即：从固定的输入推断输出的内容。从脑皮层接线图可以看出，大脑活动从最基本的感觉区域——图像、声音和触觉进入的区域开始，传导至更高级的主管分析、计划和运动的神经区域，然后将指令传达给肌肉。由此可见，人是先有感觉，然后才有行为的。

我并不是说所有人都忽略了时间和反馈。在如此广阔的科学领域中，任何一种想法都有其追随者。近年来，越来越多的人认识到时间、反馈以及预测的重要性；然而人工智能和经典神经网络的蓬勃发展，使得其他研究方法在许多年中都未能引起人们的重视。

不论外行还是专家都认为，行为是智力的特征。这一点不难理解。至少在以前的几个世纪中，人们总是将大脑的功能和钟表、水泵，后来的蒸汽机，再后来的计算机联系在一起。近十几年的科幻小说中都充斥着人工智能思想，从阿西莫夫(Asimov)的机器人理论到《星球大战》中的C3PO，认为智能机器必须能做点什么的观念已经在人们的想像中根深蒂固。所有的机器，无论是建造出来还是想像出来的，都必须能做些事情；世界上没有会思考的机器，只有能干活的机器。甚至当我们在观察自己的同类时，所关注的也是他们的外在行为，而不

是内在思想。因此，我们会直觉地认为，智能行为显然应该是智能系统的衡量标准。然而，综观整个科学史，我们不难发现直觉往往是发现真理的最大障碍。科学的框架体系很难被发现，不是因为它们复杂，而是因为直觉上的错误假设使我们无法找到正确的答案。哥白尼(Copernicus, 1473—1543)以前的天文学家错误地假定地球处在宇宙的中心固定不动，原因就是人们感觉它是不动的，而且看上去是在宇宙的中心；人们凭直觉可以清楚地看到；星星形成了一个围绕地球旋转的巨大天体。如果有人说地球像陀螺一样自转，它的表面每小时移动 1 000 多千米，而且地球本身还在天空中飞驰(更不要说那些星星在距离我们上万亿千米的远方)——这足以让别人把他当作疯子。然而这一切后来被证明都是正确的。道理很容易理解，但从直觉上看却是错误的。

在达尔文之前，物种的形式显然是固定不变的。鳄鱼与蜂鸟不可能相提并论，因为它们是如此不同甚至相互对立。而现在，物种进化的思想不再与宗教教义相悖，而且已经成为人人皆知的常识。进化论认为，你和居住在这个星球上的所有生物一样，拥有共同的祖先，包括蠕虫甚至你厨房中摆放的花草植物。我们知道这是对的，但直觉却让我们有相反的感觉。

我之所以提到这些著名的例子，是因为针对智能机器的研究也受到了这种直觉猜测的限制，阻碍了我们的发展。你会问自己：“智能系统可以做什么？”从直觉上看，它显然是通过行为来思考，我们通过演讲、写作和各种动作来展示我们的智慧，难道不是吗？是的，但不完全是。智力是某种发生在我们头脑中的东西，行为只是它的一部分。这一点从直觉上看不明

显，但理解起来也不困难。

1986 年春天，我阅读了大量科学论文，整理了智能研究的历史，并时刻关注着人工智能和神经网络领域的不断变化。整日的伏案苦读之后，我发现自己被细节淹没了。还有源源不断的材料等着我去研究，但对于整个大脑是如何工作的或者说它到底做了什么，我仍然无法找到一个清楚的答案，原因是神经科学领域本身就充斥着细节，直到现在仍然如此。尽管每年都会发表上千篇研究报告，但这些报告只是徒增书桌的负担而已，对于这些细节的组织毫无益处。对于大脑的作用以及工作原理仍然没有形成全面的理论和框架。

从此，我开始思考解决这个问题的方法。它会不会因为大脑的复杂而极为复杂呢？是否要用写满 100 页纸的数学公式才能描述清楚大脑是如何工作的？我们是否要画出成百上千的线路图才能做出一个有意义的解释呢？我不这样认为。从历史上看，对于科学问题最完满的解决办法往往是简单而精确的。尽管细节工作可能令人生畏，通往最后理论的道路可能充满险阻，但最终的概念框架体系通常是简明的。

因为没有有一个核心的解释作为引导，神经科学家无法将它们收集到的所有细节统一起来。大脑是由无数细胞构成的，数量之多、结构之复杂令人生畏。乍看来，它就像一个装满煮熟的意大利面条的体育场，被称为是机械师的噩梦。但如果仔细观察，你就会发现大脑并不是混乱无规律的，它也有自己的组织和结构，只不过它们过于复杂，以至于我们无法凭直觉想像

出它们是如何作为一个整体工作的，就像我们不敢想像如何将打碎的花瓶碎片重新拼凑起来一样。之所以不能想像，不是因为我没有足够的数据或没有足够的正确数据，我们需要的是换一个思考的角度。有了合适的大框架，细节才会有意义、可操控。看看下面这个有趣的比喻，你就能明白我的意思。

假想一下，1 000 年后人类已经灭绝，外星高级文明生物登陆地球，他们想搞清楚我们是如何生活的，尤其对我们的公路网络感到疑惑不解。这些古怪而复杂的结构是干什么用的？于是他们地面探查，卫星拍照，整理各种资料。他们像一丝不苟的考古学家那样，从每一个由沥青制成的零星碎片的位置，到自然侵蚀形成的高沟低坎上的路标，记录下所有可以找到的细节。慢慢地他们注意到有些道路网络与其他的不同：有些地方道路崎岖、狭窄，看上去好像毫无规划；而有些地方的道路却形成规则的漂亮方格；延伸一段之后，道路又会变得稠密起来，或是绵延上千千米穿越整个沙漠。他们收集到了堆积如山的详细资料，但这些资料不能说明任何问题。他们只得继续寻找，希望能找到新的线索圆满地解释这一切。就这样，他们在很长一段时间里艰难前行。

直到有一天，其中的一个家伙说：“我发现了！我想我明白了——这些生物不能像我们一样有心灵运输能力。他们只能通过一种设计巧妙的运动平台，从一个地方移动到另一个地方。”基于这个想法，许多细节都得以对号入座了：弯弯曲曲的小街道是很久以前修建的，那时的运输工具很慢；长长的密集道路是用来高速运输、长距离旅行的，这也最终解释了为什么这些道路上有不同的数字标志。由此科学家推断出哪里是居

民区，哪里是工业区，以及经贸需求和运输基础设施相互作用的方式，等等。一时间，他们整理出的许多细节变得毫不相干了，只是历史的偶然或当地地理状况要求的结果。但此时，面对同样一堆原始数据，外星生物已经不再感到迷惑了。

我可以充满信心地说，我们恰恰需要这样的突破来帮助我们分析有关大脑的所有细节。

令人遗憾的是，并不是所有人都相信我们能理解大脑的工作原理。有相当多的人都认为，从某个角度来看，大脑和智能是无法解释的，其中还包括一些神经科学家；甚至有人认为，即使我们了解了大脑和智能，也不可能建造出与之工作原理相同的机器，因为智能必须以人体神经元，甚至某些新的、高深莫测的物理定律作为基础。每每听到这些议论，总会令我想起过去的一些知识分子，他们反对研究天空，也反对通过解剖尸体研究人体的工作原理。“用不着这样做，没有任何意义。即使你弄明白了它是如何工作的，对于整个知识体系也没有任何益处。”类似这样的争辩形成了一个哲学分支——功能主义学派，它是在我们为期不长的思维研究历史上最新遇到的障碍。功能主义认为：具有智能或思维只是生物体的特征而已，与你身体的组织构成没有必然联系。思维存在于任何系统之中，只要这个系统的组成部分之间有正确的因果关系，而这些部分可以是神经元、硅芯片或者其他任何东西。很显然，这个观点是那些有可能制造出智能机器的人争论的议题。

试想，如果用一个盐瓶代替棋盘上弄丢了的马，难道这盘

棋就会因此而失去其真实性吗?当然不是。盐瓶在棋盘上来回移动,和其他棋子相互配合,它和一个真正的马在功能上是相同的,因此这盘棋也应该是一局真正的比赛,而不是模拟比赛。另外,如果我用光标将句子中的字符删掉,然后再重新打一遍,难道这个句子就和原来的不同吗?让我们再看一个更熟悉的例子。每过几年,你的身体细胞就会更新一遍,尽管如此,从你身体的重要部分来看,你还是你。只要一个原子在你身体分子的形成中发挥了作用,它就和其他原子同等重要。这种说法同样适用于大脑:如果一个科学狂人将你的每一根神经元用同样功能的微型电脑复制品替代,替代的过程完成之后,你会感到自己还是原来的自己,没有任何变化。

根据这一原理,如果一个人造系统使用了与大脑相同的智能结构,它就可以和大脑一样聪明,并且这不是刻意设计出来的,而是“真正”地具有智能。

我和人工智能的拥护者以及连接主义者都是功能主义者,因为我们相信大脑具有智能并不是出于某种天生的特殊神秘物质,总有一天我们会用某种方式造出智能机器。然而,我们对于功能主义却有着不同的解释。虽然我曾提到输入/输出谬误是人工智能模式和连接主义模式的样本失败的主要因素,并且对这一谬误进行了阐述,但我觉得还是有必要进一步说明为什么我们还不具备设计出智能机器的能力。人工智能的拥护者会认为我走的是一条自我否定的死胡同,然而在我看来,连接主义者的态度一直都过于怯懦了。

人工智能研究人员会有这样的疑问:“我们这些工程师为什么要被进化过程中偶然发现的方法所束缚呢?”从原则上看,

他们的问题不无道理。生物系统，如大脑和基因组，都被认为是粗鄙不堪、华而不实的无用品。软件设计员也有一个相应的术语——“拼凑件(kludge)”，指那些缺乏前瞻性而又极其复杂的程序。这种程序毫无用处，只会给人造成负担，有时甚至连编写者自己都看不懂。人工智能研究人员担心，大脑也会是这样一个进化了几亿年的拼凑件，一个充斥着低效能和进化“遗传密码”的混乱体。如果事实如此，他们就会问：为什么不能把这个令人失望的混乱体完全抛开，重新开始呢？

许多哲学家和认知心理学家都赞同这个观点。他们将思维比喻为大脑运行的软件，大脑和计算机硬件具有相似的功能。在计算机中，硬件和软件是完全不同的等级。同一个软件可以在任何一台万能图灵机上运行，比如，个人电脑、苹果机或超级计算机有不同的硬件配置，但它们都可以运行 Word 软件。但如果你要学习如何使用这一软件，它们的硬件系统不能提供任何帮助。通过这个类比可以看出，人类的思想自然流淌，而大脑却无法帮助我们解读思维。

人工智能的倡导者们还会提出一些历史上声动的实例，来证明工程学上的解决方法与自然之道截然不同。例如：我们是如何制造飞行器的？是模仿有翼类动物扇动翅膀的动作吗？不是。我们的飞机有固定的机翼和螺旋桨，后来又使用了喷气发动机，这显然与自然的安排完全不同，但它很有效——甚至比上下扇动的翅膀更出色。

同样，我们想造出像猎豹一样飞奔的机器，但并没有模仿它的四条腿，而是通过发明轮子，造出了比猎豹还快的陆地运输工具。轮子是在平坦陆地上移动的最好办法，而且这个特殊

的方法并不因为不是自然进化的结果而失去它的精彩。一些研究思维的哲学家喜欢用“认知轮子”来比喻对某个问题的人工智能解决办法，尽管这个办法与大脑的工作原理大不相同，但同样出色。也就是说，如果使用某个有局限但有效的程序能使某个任务的输出结果接近(甚至超过)人类的表现，那么这个程序就和大脑的功能不相上下了。

我相信正是这种“只求结果，不问手段”的功能主义解释将人工智能研究者引入歧途。正如塞尔在“中文屋”实验中表明的，行为的等同是远远不够的。智能是大脑的一个内部特征，因此我们必须研究大脑的内部以探究什么是智能。在观察大脑，尤其是新大脑皮层的过程中，我们必须仔细甄别什么是以前进化过程中形成的多余的“冻结的偶然事件”，并且肯定在大脑的重要特性之中混杂有许多无用之物。但正如我们将要看到的，神经回路之中一定潜藏有某种精确的巨大能量等待我们去发掘，而这种能量将超过任何最先进的计算机。

连接主义者凭直觉感到，大脑与计算机截然不同，而且大脑的秘密存在于相互连接的神经元的动作机理之中。这是一个好的开端，但可惜的是，这一领域到此就停滞不前了。尽管曾有上千人致力于三层神经网络的研究，甚至有些人至今仍在努力，但对于仿大脑的神经网络的研究却一直乏人问津。

半个世纪以来，我们调动了人类的全部聪明才智致力于计算机智能化的尝试。在此过程中，我们发明了文字处理器、数据库、电子游戏、互联网和移动电话，甚至还有逼真的恐龙电脑动画，但智能机器却仍然不见踪影。要想取得成功，我们必须从天然的智能引擎——新大脑皮层开始探索，必须从大脑的内部提取智能。除此之外，别无他途。

第二章

人

The Human Brain

脑



为什么大脑与人工智能及神经网络的程序如此不同?是什么使大脑的结构如此独特?而这种独特性又意味着什么?在下面几章我们可以看到,大脑的构造为以上问题提供了答案,它告诉我们大脑到底是如何工作的,为什么与计算机如此不同。

我们来看看这个器官。假设大脑放在桌子上,我们正在解剖它。首先引起我们注意的是,大脑在外表上几乎完全一致——粉灰色,就像一个表面光滑的花椰菜,中间夹杂着无数的高脊和沟壑,被称为“脑回”和“回间沟”。那个摸上去软软的、黏糊糊的部分就是新大脑皮层。它是一层薄薄的神经组织,紧紧地将大脑的其他部分包裹起来。我们的重点会集中在新大脑皮层上,因为我们所知的与智能有关的内容,如感知、语言、想像力、数学、艺术、音乐以及筹划等等,都发生在这里。此时此刻,是你的新大脑皮层在阅读这本书。

我必须承认我是一个新大脑皮层至上主义者。我知道会有许多人反对我的观点。那么,请给我1分钟,趁着还没有离题太远,让我解释一下。对于大脑的每一部分都有专门研究的科学家,如果我说通过解读新大脑皮层就可以了解智能的本质,一定会激起反对者的一片抗议之声。他们会说:“如果不首先理解大脑中某某区域的话,你是不大可能弄明白新大脑皮层的,因为它与新大脑皮层是如此密不可分,我们需要这个区域来完成某某工作。”诚然,对此我并无异议。大脑包括多个部分,其中绝大部分对人类来说至关重要。(奇怪的是,小脑是个例外。虽然它的神经细胞最多,但如果你天生就没有小脑或小脑受到损伤,你仍然可以过基本正常的生活。而大脑的其他部位却不是这样,它们是生命基本感知的必备。)

与此相对的立场是，我对造人并不感兴趣，我想做的是了解智能和建造智能机器。造一个人和拥有智能完全是两码事儿。一台智能机器不需要有性欲、饥饿感，也不必有脉搏、肌肉、情感或与人一样的躯体。人不仅仅是一台智能机器，我们既具备生物体所有必需的条件，也不乏千百年来进化过程中留下的冗余之物。如果你想制造一台模仿人类行为的智能机器，也就是说，在各个方面超越图灵机的话，你必须创造出那些作为人所必需的其他因素。下面你将看到，如果要制造出一台与人不完全相同的智能机器，我们只需关注大脑中与智能有关的部位即可。

有人会对这个看法心存异议，而我要说明的是，我也认为大脑的其他部位，如脑干、基底神经节(basal ganglia)和杏仁核(amygdala)等对于人类新大脑皮层的功能至关重要，这一点毫无疑问；但我希望你能够接受这样的观点：所有的智能都产生于新大脑皮层，并且另外两个部位——丘脑和海马，也发挥了重要作用。对于这两个部位，我会在以后作进一步讨论，因为从长远来看，我们需要了解大脑各个部分的功能和作用，但我相信这一切将会在完整的新大脑皮层理论的基础上得以完美地解决。这就是我的两点看法。现在，让我们重新回来审视新大脑皮层，或是简称它为大脑皮层。

拿出 6 张名片或扑克牌，将它们垒成一叠，作为一个大脑皮层的模型。(这样做比凭空想像要好得多。)6 张名片大约有 2 毫米厚，可以帮助我们感觉一下大脑皮层的厚度。大脑皮层像你手中的名片一样，也是 2 毫米厚，分为 6 层，每一层的厚度大约相当于 1 张名片。

如果完全展开，人类的大脑皮层大约相当于一张大餐巾的大小。其他哺乳动物的要小一些：老鼠的有1张邮票大；猴子的大约相当于一个商业信封。撇开大小不说，它们的大脑皮层也像你手中的名片一样分为6层。人类之所以更聪明，是因为我们的大脑皮层相对于体型来说更大，而不是因为它更厚或含有某种特殊的“聪明”细胞。这个大小相当关键，因为它正好将大脑的大部分包裹起来。为了和这个较大的大脑相匹配，大自然必须改变我们人体的结构，于是女人就有了一个可以分娩出大头婴儿的骨盆。这一特征被古人类学家认为是与两腿直立行走能力共同进化的结果。然而这还不够，大脑皮层又进一步进化成折叠的状态，被塞进颅骨里，就像一团塞进白兰地酒杯的纸团。

你的大脑皮层中遍布神经细胞，也叫神经元，它们紧紧排列，没有人能说出准确的数目。在你手中的一沓名片上画一个边长1毫米的方块，就可以找出大约10万个神经元所处的位置。你能数出这么小的方框中的数字吗？几乎不可能。但仍然有一些解剖学家估计人类大脑皮层中包含大约300亿个神经元，但我觉得，即使这一数字再大上很多或小上很多，也不会有人对此表示惊讶。

那300亿个细胞就组成了你的大脑。它们构成了你的记忆、知识、技能，并帮你积累生活经验。尽管对大脑研究了25年，我仍然对此惊讶不已。是这样一层薄薄的细胞让我们去看、去感受，并形成对世界的看法，这真是不可思议。夏日的温暖和对美好世界的憧憬都是由这些细胞创造出来的。曾经在《科学美国人》发表文章的弗朗西斯·克瑞克在许多年后又写了一本关于大脑的书——《惊人的假设》(The Astonishing Hypothesis)。



所谓惊人的假设，是指我们的思想就是大脑细胞的产物，没有魔力，也没有特殊的浆汁，思想就是由神经元和闪动的信息流构成的。我真希望你们能够感受到这一点是多么奇妙。尽管在哲学范畴中，细胞群和意识经验之间存在着巨大的差异，但思想和大脑却是统一的。克瑞克称之为假说，他是正确的。事实上，思想是由大脑细胞创造出来，这并不是假说。我们需要弄清楚的是这 300 亿个细胞到底做了什么、是如何做的。所幸的是，大脑皮层并不是一团乱七八糟的细胞，我们可以深入研究它的结构，以期发现它是如何创造出人类思想的。

让我们回到解剖台，再仔细看看大脑这个器官。用肉眼观察，大脑皮层上几乎没有任何标记；更确切地说，只有一个将大脑分为左、右两个半球的裂纹和一个将大脑分成前、后两个区域的回间沟。无论从左向右，还是从后向前，这个回旋的表面看上去十分相似，没有明显的界限，也没有任何颜色印记区分哪里是感觉信息区域、哪里是思维区域。

然而，人们早在很久以前就知道，它的内部是有区域界限的，甚至在神经科学家认识到任何有助于大脑皮层回路发挥作用的東西之前，他们就意识到某些心理功能是固定在一定的区域里的。如果脑中风损坏了某人的右脑顶叶，他就会失去知觉，甚至整个左侧身体失去知觉，对身体左侧的空间没有任何感觉。如果中风发生在左前脑的布洛卡(Broca)区，他就会丧失使用语法的功能，而同时他的词汇和理解词汇的能力却不会有任何改变。如果纺锤状脑回发生中风，他就不再拥有辨别面

孔的能力——就不能认出自己的妈妈和孩子，甚至连照片上自己的面孔也会变得陌生。这些奇特的功能紊乱，使早期的神经学家意识到大脑皮层包含许多不同的功能区。

上个世纪，我们对大脑功能区的了解取得了很大进展，但仍然存在许多未解之谜。每一个功能区都是半独立的，主管感觉和思维的某一个方面。这些区域的分布就像百纳被上的图案，毫无规则可言，但是每个人之间的差别却是微乎其微的。这些功能之间没有明确的界限，是按照树枝状等级排列的。

认识这种层级结构至关重要，因此我想多花些时间对它做一解释，而且我会在整本书中多次提及它。在一个层级结构中，某些元素凌驾于其他元素之上，或是屈居其下，这个概念是绝对的。在职务阶层系统中，邮局中层经理居于办事员之上、副总裁之下，这与身体所处的位置没有任何关系，纵使经理所在的楼层比办事员的低，从阶层上看，他仍然高于办事员。我强调这一点是为了帮助大家理解大脑的某个功能区与另一个功能区之间高低阶层的含义。这种阶层的高低和它们在大脑上所处的位置无关。大脑皮层的所有功能区都分布在回旋的脑皮层上，决定它们之间“高与低”的关键在于相互之间的连接。在大脑皮层中，低级区域通过某种特定的神经连接方式向高级区域传输信息；而高级区域用另一种方式向低级区域反馈信息。处于不同支系的各个区域之间还存在着连接，就像是中层经理与另一地区同级别的办公室的经理存在联系一样。科学家丹尼尔·福尔曼(Daniel Felleman)和大卫·范艾森(David van Essen)已经绘制出了详细的猴脑皮层图，从图上可以看出许多区域在一个复杂的层级结构中相互联系。由此我们可以假设，

人类大脑皮层也有这样类似的层级结构。

低级感知区是最低的功能区，也是感知信息最先到达的大脑皮层区域，它所处理的是最原始、最基本的信息。例如：视觉通过被称为 V1 区的主感知区进入大脑皮层，V1 区感觉到的是最低级的视觉特性，如尖状物、运动的小幅变化、双眼之间的差异(对立体影像而言)、基本的颜色以及对比信息。V1 区将这些特性向上传给 V4 区和 IT 区(对此我会在后面谈到)以及其他一系列的区域。每一个区域都与这些信息中的某个更专业化、更抽象的方面相联系。例如：V4 区中的细胞会对中等复杂的物体，如红色或蓝色的星状物体等做出反应；另一个被称为 MT 区的区域，专门与物体的运动发生关系。那些显示各种物体，如面孔、动物、工具、身体部位等的视觉记忆区，处在大脑视觉皮层的较高层级。

你的其他感觉也有相似的层级结构。大脑皮层上有一个被称为 A1 区的低级听觉区和更高级的听觉层级结构，以及一个被称为 S1 区的低级体觉(身体的感觉)区和更高级的体觉层级结构。某些大脑皮层区域，被称为“联合区”，它们可以接收来自多种感觉的输入信息，如来自视觉和触觉的输入信息；最终，感觉信息会传入“联合区”。正因为有了“联合区”，你才能将痒痒的感觉和手臂上爬动的苍蝇联系在一起。大部分这样的区域接收到的是由各种感觉传来的经过处理的输入信息，但他们的功能如何仍是一个谜。以后我还会进一步探讨大脑皮层的层级结构。

大脑前叶还有几个区域专管运动输出。大脑皮层的运动系统也有层级之分，其中最低级的 M1 区向脊髓传送指令并直接

控制肌肉的活动，高级区域向 M1 区传输复杂的运动命令。运动区和感觉区的层级结构非常相似，似乎是用同样的方式组成的。一般我们认为，运动区的信息是从层级结构向下传输到 M1 区并驱动肌肉的，而感觉区的信息是向上传到层级结构的。然而事实上，这些信息的传输方向是一致的——在感觉区看来反馈的就是运动区的输出，反之亦然。

大多数对大脑的描述都以流程图作为基础，而这些流程图反映出来的对层级结构的看法极其简单。它们认为，输入信息（如影像、声音和触觉）从低级感觉区进入，经过处理后传输到高级的层级结构，然后又经过联合区，进入脑皮层前叶，最后传回到运动区。我并不认为这种看法完全错误。当你大声朗读时，视觉信息进入到 V1 区，并传入联合区；然后，向前部的运动皮层传输，致使你的嘴部、喉部的肌肉发出声音。但这并不全面，事实没有这么简单。在这种过于简单化的看法中，我要提出的反对意见是，整个过程通常被认为信息好像是单向传输的，就像在工厂单向传动的流水线上装配一个小装置。事实上，大脑皮层的信息也会反向传输，产生的向下投射要比向上的多。当你朗读时，高级区域向低级的视觉皮层传输的信号要比你的眼睛从书本上接收到的信号多得多。在以后的章节里，我会谈到这些反馈投射的作用，但现在，我希望大家能注意到这个事实：尽管向上的层级结构是真实的，我们同时也必须避免这样的想法，即所有的信息流都是单向的。

再回到解剖台。假设我们有一台功能强大的显微镜，切下一层薄薄的大脑皮层，给细胞着色后，通过显微镜观察它。如果我们将切片的所有细胞都着色的话，因为细胞排列相当紧密，

我们会看到黑糊糊的一团。但如果用着色剂给一小部分细胞染色的话，我们就会看到上面提到的大脑皮层的6个层次，每一个层次的细胞密度、细胞形态和它们之间的连接都是有变化的。

神经元拥有共同的特性。从形态上看，它们正如你想像的那样是圆形的；除此之外，还有一些树枝状的结构，称为轴突和树突。当一个细胞的轴突接触到另一个细胞的树突时，就会形成一些连接，叫做“神经突触(synapsis)”。一个细胞产生的神经冲动对另一个细胞的行为造成的影响就发生在这里。当一个动作电位(action potentials)到达突触后，接受细胞很有可能产生动作电位；但一些突触也有相反的作用，反而使得接收细胞产生动作电位的可能性变小。突触的强度可以根据两个细胞的行为而发生变化。当两个神经元几乎同时产生动作电位时，它们之间的连接力度就会被加强，这是最简单的突触强度变化，被称为海布学习法则(Hebbian learning)。对此我会进一步讨论。除了突触强度的变化之外，有证据表明，两个新的神经元之间可以形成全新的突触。这种现象随时都会发生，但科学家的证据却存在争议。撇开这一点不谈，有一点是确定无疑的，即突触的形成和强度正是促成记忆存储的关键。

大脑皮层由众多神经元构成，其中最大的一类是金字塔形神经元，顾名思义，它的形状近似金字塔，占到总细胞数的4/5。大脑皮层6层结构中，最上层有长达几千米的轴突区几乎没有细胞。除了这一层以外，其他的五层中都有金字塔形神经元，它们与离自己最近的其他神经元相连，并伸出长长的轴突，向旁边的大脑皮层区延伸，甚至可到达较低的大脑结构——丘脑。

一个标准的金字塔形细胞有几个突触，它们的密度极大

且个体极小，因而确切的数字不得而知，而且具体数字也因所在细胞、皮层及区域的不同而有所变化。如果保守地估算一下，一个普通金字塔细胞有 1 000 个突触(实际上这个数字应该接近 5 000 ~ 10 000)，那么大脑皮层中就会有大约 30 兆个突触，这个天文数字远远超出了我们直觉可以理解的范围。显然，这足够你用来存储一辈子学到的东西。

据传说，阿尔伯特·爱因斯坦(Albert Einstein)曾经说过，广义相对论的理论构想并不复杂，简直可以说是易如反掌。下面的观察结果能够自然得出这个结论，即：纵使观察者以不同的速度移动，对他们来说光速都是不变的。这与我们的直觉是相反的，就好像在说，一个人投球，无论他用多大的力气，无论投球的速度和对球的观察有多快，被投出的球的运动速度都是一样的。无论在任何情况下，任何人看到的球的飞行速度相对于他们来说都是相同的。事实似乎不是这样。但对光来说，这一点被证明是正确的。就此，爱因斯坦很巧妙地提出了一个疑问：这一怪诞事实所产生的结果到底是什么？对于恒定光速的意义，经过周密的思考之后，他得出了一个关于广义相对性的更加怪诞的预言，即：当你的运动速度加快时，时间就会变慢，此时能量和质量基本成为一体。各种有关相对论的书籍引用了火车、子弹、闪光灯等生活事例粗略地解释了这种推理方式，其理论并不复杂，但却与直觉相反。

神经科学方面有一个类似的发现，这一有关大脑皮层的事实如此惊人，以至于一些神经科学家不相信其真实性，而另外

一些人却干脆置之不理，因为他们不知这到底有什么用处。然而它却是一个极其重要的事实。认真研究它的含义，如果方法得当，就可以帮助我们揭示大脑皮层的作用及其工作原理。这个惊人的事实存在于大脑皮层本身的结构中，但却需要与众不同的、富有洞察力的头脑去发现它。拥有这种头脑的人就是位于巴尔的摩的约翰霍普金斯大学的神经科学家弗农·蒙卡斯尔(Vernon Mountcastle)。1978年，他发表一篇题为《大脑功能的组织原则》(An Organizing Principle for Cerebral Function)的论文。文章指出，大脑皮层在外表和结构上惊人地相似，不论是主管视觉输入和主管触觉的大脑皮层区域、控制肌肉的区域、布洛卡语言区以及其他各个区域，实际上是完全一样的。他还暗示说，既然这些区域是相同的，那么它们在实际中所发挥的基本作用也可能是相同的，而且大脑皮层完成各个功能所使用的方法也是相同的。

在蒙卡斯尔时期以及之前的几十年中，解剖学家就已经意识到大脑皮层的各个部分看上去十分相似，这是毋庸置疑的事实。但那时他们并没有深究此项发现的意义，而是花费大量的时间去寻找它们之间的不同之处，而且也确实找到了。他们想当然地认为：如果一个区域主管语言，另一个区域主管视觉，那么它们之间肯定存在差异，只要你努力，就一定能找到这些差异；大脑皮层的各个区域在厚度、细胞密度、细胞形态的相对比例、横向连接的长度以及突触密度等方面都有所不同，而且很难被发现。低级视觉区 V1 区是他们研究最多的区域之一，其中一层就存在着一些额外的分界线。这一状况与 19 世纪中期的生物学家们面临的工作有几分类似：他们花费大量时间

来寻找物种之间的细微差别，结果唯一的发现是两种外表几乎相同的老鼠属于不同物种。许多年中，达尔文在研究软体动物时也选择了同样的道路，但他最终醒悟了，并发出了这样的感叹：所有这些物种竟然如此相似！真正令人惊讶和感兴趣的不是物种们之间的差别，而是它们的相似性。

蒙卡斯尔也做了类似的观察。在解剖学家们找寻大脑功能区域的细微差别之时，他证明了大脑皮层尽管有差异，其相似之处才是惊人的。就像你手中的6张名片一样，它们的层数相同、细胞种类相同，甚至连接也相同，其中的差异之细微连神经学家也无法对此达成共识。因此，蒙卡斯尔说，大脑皮层所有的区域功能相同，而视觉区域之所以能“看见”、运动区域之所以能使肌肉运动，都是由这些区域之间或与中央神经系统的其他部分之间的连接决定的。

实际上，蒙卡斯尔所说的大脑皮层区域之间的微小差异正是它们之间连接的差异，而不是基本功能的差异。他总结说，所有大脑皮层的功能区域都遵循一个共同的算法，视觉、听觉、甚至运动输出之间没有任何差异。他还认为，大脑皮层的连接方式是由基因决定的，这正是功能和物种的独特之处，而脑皮层组织本身在各个区域都担负着相同的作用。

让我们再斟酌一下。我感觉视觉、听觉和触觉是完全不同的，它们有着迥异的基本特性。视觉跟颜色、质地、形状、深度以及轮廓有关；听觉则包括音高、节奏和音质。它们看上去如此风马牛不相及，怎么可能一样呢？蒙卡斯尔说，它们不一样，但大脑皮层处理从耳朵和眼睛传来的信号时，方法是相同的，而且，运动控制系统的工作原理也是同样的。

大多数科学家和工程师都忽略了，或者说故意忽略了蒙卡斯尔的观点。他们想研究视觉，让计算机也能“看见”，但他们只是在一些特殊的视觉词汇和技术上下功夫，满口“边缘”“质地”“三维立体显示”等专业名词。如果他们想要研究口语，就会建立一个以语法规则、句法和语义学为基础的规则系统。但如果蒙卡斯尔的观点是正确的，这些方式都将失败，因为它们与大脑解决问题的方式完全不同。如果蒙卡斯尔的观点是正确的，大脑皮层的算法必须独立于其他特殊功能和感知。大脑用同样的方式去“看”、去“听”，大脑皮层的作用具有普遍性，适用于任何一种感觉或运动系统。

当我第一次读到蒙卡斯尔的论文时，激动得差点从椅子上跌下来。它正是帮助我们解读神经科学的罗塞塔石碑(罗塞塔石碑是 1799 年在埃及的罗塞塔镇附近发现的古埃及石碑，其碑文用古埃及象形文字和通俗文字以及希腊文字刻成；该石碑的发现为解读古埃及象形文字提供了很大帮助。罗塞塔石碑用以指那些有助于理解疑难事物的事物。——译者注)——仅仅一篇论文和一个观点就将有关人脑的形形色色的奇特功能连接了起来，而且它们的算法是统一的。它轻而易举地证明了，以往那种将人类行为作为不同功能进行理解和构想的尝试是错误的。我多么希望大家能体会到蒙卡斯尔观点的简洁与彻底。最棒的科学观点往往是简洁、彻底且不同凡响，而这一观点就是其中之一。我认为它曾是，也将一直是神经科学界最重要的一个发现。然而，令人难以置信的是，大多数的科学家和工程师要么是拒绝相信它，要么是置之不理。

这种漠视的态度部分原因是缺乏研究工具。我们没有合适的仪器来探究信息是如何在 6 层大脑皮层中传播的，现有的仪器只能进行一些粗糙的试验，通常是以确定各种功能在大脑皮层中出现的位置为试验目标，而不是确定其时间和方式。例如，最近在大众媒体上发表的神经科学论文就明确支持这样的观点，认为大脑是一个高度专业化的单元集合。像核磁共振成像(MRI)和全身骨扫描显像(PET)等功能造影技术，无一例外地将研究的重点集中在前面提到过的大脑功能区和分布图上。在这些试验中，志愿受试者将头伸进扫描仪，躺着完成一些心理活动和运动任务，如玩电子游戏、进行动词变位、读句子、看不同的脸孔、指认图片、想像、记忆一连串的东西、做财务判定等等。扫描仪可以监测出完成这些任务时大脑的哪些部分更加活跃，并在受试者的大脑影像图上将这些区域涂上颜色标示出来，由此就可以推断出这些区域就是任务的中心。这种功能造影实验做了上千次，以后还会再做上千次。由此，我们逐步对于某些功能在正常成人脑中发生的位置有了了解，并可以毫不费力地说出：这是面孔辨认区，这是数学区，这是音乐区，等等。我们不知道大脑是如何完成这些工作的，因此就想当然地推测出，大脑在用不同方式进行各种活动。

事实的确如此吗？有越来越多的有力证据支持蒙卡斯尔的理论，一些最好的事例也证明了大脑皮层具有极强的灵活性和可塑性。如果营养合理、所处环境良好，任何一个人脑可以学会几千种口语中的任何一种，同时还能学会符号语言、书面语

言、音乐语言、数学语言、计算机语言和肢体语言，等等。它能学会适应北方地区的寒冷，也能在炎热的沙漠地区生活。它可以成为象棋高手，钓鱼或种庄稼的好手，也能成为理论物理的专家。试想，你大脑中小小的视觉区域似乎专门处理手写的字母和数字，难道这就意味着你的大脑天生就具有处理文字和数字的语言区吗？不可能！书面语言是我们的基因最新形成的一种特别进化机制，因此，早在孩提时期，大脑皮层就分化成不同的特殊任务功能区域，其分化基础是经验。人脑学习和适应不同环境的能力之强令人难以置信，而这些能力是在最近的时期才形成的。这一点就证明它极具灵活性，而不是所谓问题与解决方法一一对应的系统。

神经科学家还发现，大脑皮层的回路具有令人惊异的可塑性，也就是说，它可以根据流经的输入信息的类型进行改变和重组。例如，刚出生的雪貂的大脑经过手术重新连接后，它的眼睛可以将信号传输到本应是听觉发育的大脑皮层区。其结果令人惊讶，雪貂大脑中的听觉区就形成了视觉传输途径，也就是说，它们看东西时使用的大脑组织本应是用来听声音的。科学家也对其他知觉和大脑区域做了类似的试验。例如，在老鼠刚出生时将几块视觉大脑皮层移植到它的触觉区域，当它发育成熟后，被移植的组织处理的是触觉而不是视觉信息。可见，细胞在形成之初并没有专门从事触觉、视觉或听觉的区分。

人类大脑也具有同样的可塑性。先天耳聋的成年人，视觉信号的处理通常发生在处理听觉信息的区域。先天性盲人在阅读布莱叶(braille)盲文时，通常使用大脑皮层后部的一个区域，而这个区域通常是主管视觉的。阅读布莱叶盲文时需要触觉，

你可能会认为它主要激活的是触觉区——但很明显，任何大脑皮层区域都不会无所作为。当视觉区没能收到“本应”从眼睛传来的信息，就会设法从其他皮层区域寻找其他形式的输入并使其通过。

这一切都说明，大脑各区域是根据传入的信息种类而发展出专门的功能。大脑皮层并不是严格地运用不同的算法完成不同的功能的，正如目前地球上各个国家区域的划分并不是先天注定的。和地球上的政治地理一样，假如在早期设定一个不同的环境，你的大脑皮层也可能会与今天的情况完全不同。

基因决定了大脑皮层的整个建构，包括各个区域之间相互连接的具体细节，但在此结构的内部，系统却具有极大的灵活性。

蒙卡斯尔是正确的。不同的大脑皮层区域有着一个相同的、强大的通用算法，如果将这些区域按照合适的层级结构连接起来并输入信息流，它就能学会了解周围的环境。因此，未来的智能机器不必具有和我们人类相同的直觉和能力。如果在一张人造机器皮层上用新的方法将大脑皮层的算法调动起来，使其具有新的知觉，那么，具有灵活性的真正智能就出现了，它不依赖我们的生物大脑。

下面要谈到的话题与蒙卡斯尔的理论有关，同样会令你惊讶不已，那就是：进入你大脑皮层的输入信息是基本相同的。你也许认为你的各种知觉是完全不同、完全独立的，毕竟，声音是通过空气传播的压缩震动波，影像以光的形式传播，而触觉是对皮肤的一种压力。声音似乎是时间的，影像主要是形象

化的，而触觉主要是空间的。怎么能将山羊的咩咩叫声和一个苹果的样子以及触摸篮球的感觉相提并论呢？还有什么比这差别更大？

让我们仔细分析一下。从外部世界传来的视觉信息通过上百万根视觉神经纤维传入大脑，快速通过丘脑之后，这些信息进入到基本视觉大脑皮层。声音是由听觉神经的 3 万根纤维传输的，经过一些较古老的大脑部位后，进入基本听觉皮层。触觉和内部感知信息是通过另外上百万根神经纤维由脊髓传输的，最终由基本体觉皮层接收。这就是你的大脑接收到的主要信息，也是你对整个世界的感知。

你可以将这些输入看作一束电线或一束光纤。你也许见过由光纤制成的灯，每一根光纤的顶端都可以发出有颜色的光。输入大脑的信息与此很相近，但它们不叫“光纤”，而被称为“轴突”；它们承载的神经信号被称为“动作电位”或“电脉冲(spikes)”，同时具有化学特性和电子特性。虽然这些信号来自不同的感觉器官，一旦转化成由大脑控制的动作电位，它们就变成了完全相同的模式。

当你看到一只狗，一组模式就通过你的神经纤维传入大脑皮层的视觉区域；当你听到狗的叫声，另一组模式就通过听觉神经进入大脑的听觉区域；当你轻轻拍打狗时，一组触觉模式就会通过你的手，经由脊柱的纤维进入大脑专管触觉的部位。看到狗、听狗叫、触摸狗——这些模式的感觉不同，因为它们进入大脑皮层层级结构的途径不同，而真正起作用的是这些信息在大脑中的传输去向。但从知觉输入的抽象层面上来看，它们都是基本相同的，都是由 6 层大脑皮层用相似的方式进行处理

的。你能听到声音、看到光、感觉到皮肤上的压力，但在大脑内部，这些信息之间并不存在根本差异。动作电位只是动作电位而已，不论源自哪里，这些瞬间的动作电位是完全相同的。你的大脑所知道的只有模式。

你对世界的认识和有关世界的知识都来自于这些模式。你的头脑中没有光，只有漆黑一片；也没有声音进入你的大脑，那里一派宁静。事实上，大脑只是你身体的一个部分，它本身没有任何感觉；外科医生可以将手指伸进去，而你却毫无察觉。你头脑中的所有信息都是作为轴突上的空间-时间模式进入的。

何为“空间-时间模式”？让我们看看人类的主要知觉：

视觉承载的是空间和时间信息，所有的空间模式与时间契合，它们形成于同一感官的不同感受器同时受到刺激之时。对视觉来说，视网膜是感觉器官。当一个影像进入瞳孔，晶状体将其翻转，投射到视网膜上，然后传递给大脑。人们一般认为进入视觉区域的是一个颠倒的小图像，而实际没有任何图像，因为所谓影像从根本上说是一些以某种模式激活的电子动作。大脑皮层对这些信息进行处理，使它们在不同区域之间的各个模式成分中传递、筛选和过滤，在这一过程中，这些信息的影像特征很快就丧失殆尽了。

视觉同样依赖于“时间模式”，即：进入眼睛后不断变化了的模式。如果说视觉的空间模式比较容易从直觉上理解的话，这种时间模式就不是很明晰。你的眼睛每秒钟会快速移动3次，它们注视一个点，然后很快又会突然跳到另一个点。随着眼睛的每一次移动，视网膜上的影像就会不断变化，也就是

说，输入你大脑的模式会随着每一次的眼扫视而彻底改变，而且这是当你目不转睛地盯着一个景物时发生的最简单的情况。在现实生活中，你会不停转动头和身体，不停地穿行于不断变化的环境中。你感觉到的是一个充满了可捕捉的事物和人物的稳定世界，但这种感觉是由大脑创造出来的，因为它有能力处理视网膜上那一连串毫无重复的影像。自然的视觉模式像一条河流进大脑，与其说它是一幅图，还不如说它像一首歌。

许多视觉研究者忽略了眼扫视和快速变化的视觉模式。他们观察被麻醉的动物，借此研究失去意识的动物在将眼睛固定在一点上时是如何产生视觉的。这样做撇开了时间量的考虑，从原则上看无可厚非，因为剔除可变量是科学研究方法的关键要素。但他们这样做忽略的是视觉中包含的一个核心成分，因为从神经科学对视觉的解读来看，时间占据了中心的位置。

说到听觉，我们总是考虑到声音的时间特性。直觉告诉我们，声音、口语和音乐都是瞬息万变的。你不可能立即听完一首歌，也不可能瞬间听完一个句子。一首歌总是存在于一段时间之内，但我们通常不会将声音看作空间模式。从某个方面来看，它与视觉的情况正好相反：时间模式一目了然，而空间模式却不甚明了。

听觉同样具有空间方面的成分。将声音转化成动作电位的是一个称为“耳蜗”的螺旋形器官，它个头很小，不透明，呈螺旋状，嵌在人体最坚固的骨头——颞骨上。对耳蜗全面解读是在半个世纪前由一位匈牙利医生——乔治·冯·贝凯西 (Georg von Beksey) 完成的。冯·贝凯西制作了一个内耳模型，发现不同的声音类型可以引起耳蜗的不同部位发生震动：高频

声音震动的是耳蜗坚硬的底部，低频音能引起耳蜗外部较柔软宽大部位的震动，而中频音震动的部位则处于耳蜗的中部。耳蜗上布满了神经元，当不同部位颤动时，那里的神经元就被激活。日常生活中，耳蜗随时会遭到多种频率的声音同时冲击。因此，每时每刻耳蜗中都会有新的空间刺激模式，每时每刻这种新的空间模式都会传递到听觉神经，然后，正如我们所知的，这种感觉信息都转化成空间-时间模式。

人们通常不会把触觉看成是时间现象。正如我们所知，这种感觉的点点滴滴都是由特定的时间、空间决定的。你可以做一个试验来判断一下：让你的朋友把手握成杯状，掌心向上，闭上眼睛。将一个诸如戒指、橡皮等普通的小东西放在他的手中让他辨别，条件是不能移动手的任何一个部位，他只能说出重量和大致的形状，除此之外再感觉不到任何线索。随后，仍然让他紧闭双眼，用手指触摸这个物体，这时他几乎可以立刻辨别出来，因为手指的移动给触觉认识赋予了时间。视网膜的中央凹和手指的相似之处，在于它们都极其敏锐。因此触觉也像一首歌，你之所以能拥有使用触觉的复杂能力，如扣衬衫、在黑暗中打开前门，完全依赖于这些持续不断地随时间变化的触觉模式。

我们教导孩子说，人类有 5 大知觉：视觉、听觉、触觉、嗅觉和味觉，而事实上，人类的知觉还要更多。视觉更像动感、颜色和亮度（即黑白对比度）这三种知觉的集合。触觉包括压力、温度、痛感和震动。我们还有一套完整的感官系统，叫做“知觉系统”，它使我们能感觉到自己的关节角度和身体的位置，没有它，我们就无法运动。除此之外，在内耳中还有前

庭系统，使我们具有平衡感觉。尽管这些知觉在发达和显著程度上不尽相同，但它们都是通过轴突以空间模式和时间模式进入大脑的。

你的大脑皮层不能直接感知外部世界，它所知道的只有输入轴突上的模式流。你的世界观，包括你对自己的认识，都是由这些模式形成的，你的大脑根本无从知晓你的生命终结于何处，这个世界又起源于哪里。研究身体影像的神经科学家发现，我们对自我的认识要比感觉灵活得多。例如给你一个小耙子，用它代替手伸出去抓东西，很快你就会感到它已经成为你身体的一部分了。大脑会改变它的预期以适应新的触觉输入模式，因此，小耙子已经在你的身体部位图上具体显现出来了。

来自于不同感觉的模式在大脑中都是相同的，这一观点确实惊人；尽管人们对它已相当了解，但该观点仍然没有被广泛接受，于是更多的实验接踵而至。下面的实验你可以在家里做，所需的只是一位朋友、一个独立的小屏风和一只假手。第一次做实验时，最好能找到一只橡胶手，就像万圣节商店里卖的那样。如果没有，也没关系，可以在一张白纸上画出你的手形。将你的手放在桌子上，和假手排成一线，中间相距 10 厘米（手指朝向要相同，手掌方向也要相同，要么都朝上，要么都朝下。）然后用屏风挡在两手之间，这时你只能看到那只假手。当你注视假手时，让你的朋友同时轻抚着两只手的相同部位，例如，用同样的速度轻轻触摸小手指上从掌关节到指甲的部位，然后以同样的节奏轻拍食指的第二个关节，接着在两个

手掌上画几个圆圈，等等。一段时间后，你大脑中的视觉和体觉模式相结合的区域（这一接合区曾在本章开头介绍过）就会混乱迷惑——你似乎能够感觉到从看到的假手传来的感觉，就好像它是你的真手一样。

另外一个模式等同实验叫做“感觉替代”。这一实验可能会给那些从童年期就失明的人的生活带来革命性的改变，也可能在不远的将来惠及那些天生的盲人，还可能为我们这些正常人带来新的机器接口技术。

认识到大脑与模式的关系之后，威斯康星州大学的生物医药工程学教授保尔·巴奇·瑞塔(Paul Bachy Rita)发明了一种在人的舌头上显示视觉模式的方法，戴上他研制的装置，盲人就可以通过舌上的感觉学会“看”。

其原理是这样的：在受试者的前额上戴上一个小型摄像机，这时视觉影像就一个像素一个像素地传输到舌头上的压力点上，一个在电视屏幕上由无数个像素组成的视觉影像可以转化成一个含有无数个压力点的模式，大脑会很快学会如何正确辨别这些模式。

第一个使用这种装在舌头上的装置之人是艾瑞克·韦恩梅(Erik Weißenmayer)。他是一位国际级运动员，13岁时失明，曾到各处讲演，鼓励人们不要让失明剥夺了自己的远大理想。2002年，韦恩梅登上了珠穆朗玛峰，成为尝试并完成此项挑战的第一位盲人。

2003年，韦恩梅试用了这个舌头装置，自童年失明后第一次看到了图像。他看到一个球在地板上向他滚过来，他伸手拿起了桌上的一杯饮料，还玩了“石头、剪刀、布”的游戏。然

后他又沿着一条走廊前进，看到了门，并察看了其中一扇门和门框，发现上面有一个标牌。原本是舌头上的感觉很快就变成了对空间影像的感觉。

这些实例又进一步证明了大脑皮层是极其灵活的，而且输入大脑的都是模式。这些模式来自哪里并不重要，只要它们在时间上以固定的方式彼此联系，大脑就能感觉到它们的存在。

如果我们接受了大脑所了解的只有模式这个观点后，这一切就不足为奇了。大脑是模式机器，用听觉或视觉来表述它的功能并非不正确，但从根本来看，模式才是实质。不论大脑皮层的各个区域的行为如何不同，其中起作用的基本皮层算法是相同的。大脑皮层并不在意这些模式是源自视觉、听觉还是其他感觉，不在意输入是来自单一的一种感官还是四种感官，无论你是通过声呐、雷达还是磁场来认识世界，你没有手却长着触角，甚至你生活在四维而不是三维空间，对这一切，大脑皮层都毫不在意。

这就意味着要具有智能，不一定必须具有某种知觉或任何知觉的特殊连接。海伦·凯勒(Helen Keller)既没有视觉也没有听觉，然而她学会了语言，成为了一名比大部分健全人都优秀的作家。她是一个缺失了两种主要知觉的智能人，然而，大脑神奇的灵活性使她能够像五种知觉俱全的人一样去认识和感知世界。

这种极大的灵活性给我们发明基于大脑研究的技术点燃了希望。当我考虑建造智能机器时，就想为什么不追随着我们所熟悉的感觉走呢？只要我们能够破译大脑皮层的算法，创立一个

模式学科，便可以将它运用到任何想使之智能化的系统上。这种基于大脑皮层特征的回路，其重要特点之一就是我们不需要巧妙地编程。重新连接的雪貂大脑中，听觉区可以变成视觉区；盲人的大脑中，视觉皮层可以找到自己另外的作用；与此相同，如果在一个系统中运行大脑皮层算法，那么它就可以根据我们输入的模式而具有智能。这需要我们花些心思设计出这一系统的普通参数，并且训练它。之后，大脑进行复杂而有创造性思维过程所涉及的几十亿个神经任务就会自然完成，就像我们在孩子身上看到的那样。

模式是智能的基本媒介，这一观点引发了一些有趣的哲学论题。当我和朋友们坐在房间里，我怎么知道他们在那儿，怎么知道他们是真的呢？当我的大脑收到的一系列模式和我以前获得的模式相匹配后，这些模式就对我认识的人做出反应，包括他们的面孔、声音、举止等各方面的情况。一直以来，我学着接受这些以某种既定方式同时发生的模式，但现在看来，它们只是一个模型，我们对世界的看法是一个建立在这些模式之上的模型。我们能否肯定这个世界是真的？这个问题似乎奇怪而可笑。一些科幻小说和电影都曾探讨过这个主题。并不是说世界上的人和物不是真的，他们是真实存在的，但我们对于世界存在的肯定是建立在模式和解读它们的方式的一致性上的。直接的感知根本不存在，我们没有“人的”感觉器官。请记住：大脑是一个漆黑而寂静的盒子，里面除了在输入纤维上像时间一样流逝的模式之外别无他物。你的世界观是由这些模式创造的。存在固然是客观的，但大脑中输入轴突束上的空间-时间模式是我们必须要经历的。

这个讨论使人们注意到一个时常困扰我们的问题——幻觉与现实的关系问题。如果你能幻想出橡胶手上的感觉，如果你能通过对舌头的刺激“看”到东西，那么你自己手上的触觉和亲眼看到的景物会不会也是这种“欺骗”的结果呢？我们能相信这个世界就是我们看到的样子吗？我认为，世界的确是以一个绝对的形式存在的，和我们感觉到的非常接近，但我们的大脑不能直接认识这个绝对的世界。

大脑认识世界要借助一系列的感觉，而这些感觉并不能发现这个绝对世界的全部。感觉形成模式传入大脑皮层，经过相同的皮层算法处理之后就形成了世界的模型。这样看来，尽管口语和书面语在感觉上完全不同，但它们被感知的方法是相同的。尽管海伦·凯勒在知觉上有很大的缺失，但她头脑中的世界模型与你我的并无差异。通过这些模式，大脑皮层创造出一个与真的世界近乎相同的模型，然后，很巧妙地将它存于记忆之中。而记忆，也就是模式进入大脑皮层后所发生的一切，我将在下一章作深入阐述。



第四章

记

Memory

忆



不论你正在读这本书，还是穿行于拥挤的街道；不论在听交响乐，还是在安抚哭闹的孩子，你的大脑都充斥着大量从感官传来的各种空间/时间模式。

世界就像一个汹涌着的不断变化模式的海洋，不停地拍打、冲击你的大脑。你是如何来理解它们的呢？模式一拥而入，通过“古脑”的各个部位，最后到达新大脑皮层。然而，它们进入大脑皮层后，又发生了什么？

工业革命之初，人们就将大脑看作一部机器，尽管他们知道大脑里并没有齿轮，但这是对它最恰当的比喻，因为信息进入大脑后，大脑这部机器就会决定身体该如何反应。计算机时代到来后，大脑又被看作另一种很特别的机器——一种会编程序的计算机。第一章中提到，人工智能的研究者们坚持认为他们的研究之所以举步维艰，唯一的原因就是，和大脑相比，计算机太小、速度太慢。他们说，现在的计算机只相当于一只蟑螂的大脑，如果我们能造出更大、更快的计算机，它一定能像人一样聪明。

这个大脑和计算机的比喻大大忽略了一个重要的问题——与电脑晶体管相比，神经元要慢得多。一个神经元将突触上传来的输入信息收集起来并将它们组合，以决定什么时候向其他神经元输出脉冲，整个过程可以在5毫秒内（大约每秒钟200次）完成并使自己复位。这看起来似乎很快，但一台现代硅芯片计算机可以在1秒钟内完成10亿次计算，也就是说，一个普通电脑运算要比正常人脑运算快500万倍。这是一个极大、极大的差距！大脑怎么可能比最快的数字计算机更快、功能更强大呢？支持电脑类似人脑观点的人会说：“没问题。大脑是一个并

行计算机，它有几十亿个细胞同时进行计算，这种并行性大大加强了生物大脑的处理能力。”

我一直认为这种说法很荒谬。一个简单的思考实验就可以证明我的观点，这个实验叫做“一百步”法则。一个人可以在不到 1 秒钟内完成不同寻常的任务。例如：我给你看一张照片，让你判断照片上是否是一只猫。如果是，你就按下按钮；如果是熊、疣猪或芜菁甘蓝，就不按按钮。对于现有的计算机来说，这个任务太难，甚至不可能完成；然而一个人却可以在半秒钟，甚至更短的时间内轻松完成。但是神经元的速度很慢，在半秒钟内，进入大脑的信息只能通过 100 个神经元长度的链条，也就是说，不论有多少神经元参与了这个过程，大脑总能在“一百步”之内“计算”出问题的答案。从光进入你的眼睛到按下按钮，这一段时间只涉及仅仅 100 个神经元长度的链条。对于计算机来说，要试图解决相同的问题则需要走“几十亿步”。100 个计算机命令就连在显示屏上移动 1 个字符都略显不足，更不要说完成任何有趣的任务了。

如果有上百万个神经元共同作用，会不会形成一个并行计算机呢？不会。大脑并联运行，并行计算机也并联运行，这是它们之间唯一的相似之处。所谓并行计算机就是将多台快速计算机并联在一起处理一些复杂的问题。比如说计算第二天的天气状况。要预报天气，必须计算出地球上许多位置点的物理状况。所有计算机可以同时计算，但每台只能负责处理一个不同的位置点。尽管有上百台，甚至上千台计算机并联工作，但对每一台机器来说必须进行几十亿甚至几百万兆次的运算才能完成任务。

我们可以想像出的最大的并行计算机也无法在“一百步”内做出任何有意义的事，不论它有多大，速度有多快。

再打个比方。假设我让你从沙漠的一头到另一头运送 100 块大石头；你一次只能搬 1 块，要走 100 万步才能穿过沙漠。你觉得只靠自己要很长时间才能完成，于是就雇了 100 名工人一起干。现在工作进展的速度提高了 100 倍，但仍需要走至少 100 万步才能穿过茫茫沙漠。雇更多的工人，哪怕是 1 000 人，也不会有更大的作用。不论你有多少雇工，要完成工作就必须花费不少于走 100 万步的时间。并行计算机的道理相同。某一个数量之后，增加更多的处理器也于事无补。一台电脑，不论它有多少处理器，不论它的运行速度有多快，都不可能在“一百步”内“计算”出一个复杂问题的答案。

那么，人脑是如何在“一百步”内完成了最大的并行电脑在 100 万甚至几十亿“步”都无法完成的工作呢？答案很简单：因为大脑不需要“计算”问题的答案，它只是从记忆中取出答案。实质上，这些答案在记忆中封存已久，只需几个步骤就可以将其取出。再慢的神经元都足以完成这件工作，而且还能建立记忆。整个大脑皮层就是一个记忆系统，根本不是计算机。

我用下面的例子来说明“计算”出答案和“用记忆”找出答案有何不同。要完成的任务是抓住一只球。有人向你投来一只球，你看到它向你飞来，并在不到 1 秒钟时间内在空中抓住它。这看起来并不太难。但当你试图给一只机器手臂编写程序

以完成这个动作时，情况就完全不同了。尽管很多研究者找到了一些办法，但似乎都不可行。工程师和计算机科学家在处理这个问题时，首先要计算球的飞行，从而确定它到达手臂时的位置。这个计算包括一系列你曾在高中物理课上学过的物理公式。然后，要调整机器手臂的各个关节，使手移动到合适的位置。这一步骤包含了更复杂的数学公式。最后，整个运行需要重复多次，因为随着球越飞越近，计算机不断获得有关它位置和轨道的最新信息。如果等到球飞到应到的位置才开始动作的话，那就太晚了。它必须在还不十分清楚球的位置的时候就开始移动，并随着球越来越近，不断地进行调整。要抓住一只球，计算机需要进行几百万次大量的数学计算。即使成功地设计出了这样的程序，“一百步”法则也告诉我们，大脑在解决这个问题时使用了不同的方法，那就是记忆。

如何用记忆抓住球呢？你的大脑内储存着抓住球所需的肌肉命令记忆（还有其他许多习得行为）。当一个球投来，会发生三件事情：首先，一看到球，合适的记忆就被自动唤起；第二，这个记忆会引发一连串的肌肉命令；最后，被取出的记忆会随着当时的特定情况，如球的实际轨道和你所在的位置，不断调整。接住球的记忆不是像程序一样编入大脑的，而是通过多年不断训练后学习并储存起来的，也不是在神经元里计算出来的。

你可能会想，“且慢，每一次接球都会有所不同。你只是说复苏的记忆会不断调整以适应每一次投掷时球的不同位置的变化，难道这不需要计算那些我们试图尽力避免的公式吗？”可能会。但大自然已经用一种与众不同的、清晰的方式解决了这

种变化的问题。在下面，我们可以看到大脑皮层有一种被称为“恒定表征”的能力，可以自动处理这种变化。想像一下你坐在水床上的情景：水床上的枕头和人都会不由自主地形成一种新的姿态。事实上，水床本身并没有计算每一个物体应该被抬升多少，只是由于水的物理特性和水床的弹性表面会自动调节。正如我们将在第五章看到的，大脑皮层的6层结构，不太严格地说，对于信息的传输具有相似的功能。

因此，不论从并行还是其他方面，大脑皮层都不像计算机，它不会计算问题的答案，而是用存储的记忆来解决问题，做出反应。

虽然计算机也有记忆，且是以硬盘驱动器和记忆芯片的形式出现的，但大脑皮层的记忆和电脑的记忆有以下四点根本区别：

- 大脑皮层可以存储模式序列；
- 大脑皮层以自-联想方式回忆起模式；
- 大脑皮层以恒定的形式储存模式；
- 大脑皮层按照层级结构储存模式。

我们将在本章讨论前三个区别。在第三章我向大家介绍过大脑皮层的层级结构，并且会在第六章继续论述它的重要性以及它是如何工作的。

下一次当你讲故事时，一定不要着急，先考虑一下应该如何一次讲述故事的一个方面，因为，不论你说得有多快，不论我听得有多快，你都无法将发生的所有事情一股脑儿地讲出

来。你必须先讲完一部分，再接着讲下一部分。之所以如此，原因之一是语言的连续性，无论是口述、笔录还是图画，讲故事的过程就是用连续的方式表达一个叙述性的事件。原因之二，故事是以序列的方式存在你的大脑中的，因此也必须以同样的顺序讲出来，你无法一下子记起整个故事。事实上，要想起不连续的复杂事件和想法几乎是不可能的。

你也许已经注意到，有些人在讲故事时不能立刻切入主题，而是啰啰嗦嗦地讲一些无关的细节。这让人很恼火，甚至想大叫：“讲正题吧！”但他们只会按照发生的时间，而不是其他方式来讲述他们的故事。

另外举一个例子，闭上眼睛想像一下你的家。你走向大门，想像门的样子；打开它，走进去。现在看看你的左边，看到了什么？再看看右边，又看到了什么？接下来去你的卧室。右边有什么？左边有什么？顶层左边的抽屉里有什么？你的淋浴室里又有什么？这一切，甚至更多，你都清清楚楚，可以回忆起每一个细节。这些记忆都存储在你的大脑皮层里。你可能会说，这些是关于家的部分记忆，但不可能一下子把它们全部记起来。它们是明显相关联的记忆，但是你无法将所有细节统统回忆起来。你对家有一个完整的记忆，但当你回想时，必须按照连续的步骤，就像你亲身经历的那样。

所有的记忆都是如此。你必须经历做事情时的时间顺序，一个模式(走近大门)唤起下一个模式(进入大门)，这个模式又会唤起下一个(穿过大厅或走上楼梯)，等等，以此类推。每一步都是你曾经历过的步骤。当然，我可以有意识地改变介绍的顺序，比如，如果想用非顺序的方式重点介绍一些东西，就可

以从地下室直接跳到二层。然而，一旦开始描述选取的房间或物品时，我又会按照一定的顺序来讲。

实际上，随机的想法是不存在的，几乎每一次回忆之后总有一个联想过程紧随其后。

你很熟悉字母表，但你却不能从后往前把它背出来，因为你平时没有倒背过。如果你想体会小孩子学习字母表的感觉，就把它倒背一遍吧，确实很难。你对字母表的记忆是一个模式序列，并不是按照任何方便、任意的顺序存储或回忆起的，就像一星期里的每一天、一年里的每一个月、你的电话号码以及其他无数的事物一样。

歌曲的记忆是一个很恰当的事物按时间顺序存储在记忆之中的例子。请回想一首你所熟悉的曲调——我喜欢《飞越彩虹》(Somewhere over the Rainbow)，当然其他歌曲也可以。你不会立刻想起整首歌，而是按顺序记起。你会从头或歌曲的主要部分开始，然后一个音符一个音符唱出来。你不可能从后往前唱，正如你无法立刻记起它的全部，因为你第一次听到《飞越彩虹》时，它是从头到尾按顺序播放的。你只能按照学歌时的顺序记起它。

这同样适用于低级感觉记忆。想想你对不同材质的触觉记忆。你的大脑皮层可以记住手中握满沙砾的感觉、手指滑过天鹅绒的感觉以及按下钢琴琴键的感觉。这些记忆像字母表和那些歌曲一样，都是以序列为基础的，只不过这些序列只有短短的1秒钟，而不是几秒或几分钟。如果你熟睡时将你的手埋进装满沙砾的桶里，当你醒来时，如果不移动手指，就不知道自己的手接触到的是什么。你对沙砾的触觉记忆是建立在皮肤

上的压力和震动感觉神经带给你的序列模式之上的。这些模式完全不同于将手埋在聚苯乙烯小颗粒和干树叶中的模式。小砾石的尖利和圆滑感形成了沙砾的序列模式，从而激活了你体觉大脑皮层中相应的记忆。

下一次洗澡时，请留心你用毛巾擦干身体的过程。我发现我在擦干身体的时候，几乎每次都是相同的程序——擦干、拍净以及相应的姿势。通过有趣的观察，我还发现我妻子淋浴之后也遵循基本上一成不变的方式。你或许也是一样。如果有一定的顺序，就试着改变它。

你可以强迫自己去改变，但必须集中注意力，因为稍一走神，你又会落入习惯的方式中去。

所有的记忆都存储在神经元之间的突触连接点上。大脑皮层中存有如此大量的信息，并且在任何时候都可以将其中的一小部分取出来，由此可见，每一次记忆提取的过程中只有少数有限的突触和神经元在起作用。当你回想你的家时，相应的一部分神经元开始变得活跃，接着它们会激活另外一些神经元，以此类推。一个成年人的大脑皮层记忆容量大得惊人，但即便如此，我们也只能一次回忆起其中的一部分，并且只能按照联想顺序来记起相关的内容。

下面是一个有意思的练习。努力回忆自己的过去——你的住处、曾经去过的地方、曾经认识的人等细节。我发现自己总是可以记起那些多年来未曾触碰过的记忆。你大脑中的突触内储存着几千个从未用过的记忆细节，在某个时间，你会记起其中一个很小的片断，而大多数的信息则存在那里，静静地等待着合适的信号去激活它们。

计算机的记忆不是按照顺序模式存储的，要想实现这种存储模式，我们也可以借助一些软件的帮助(就像你在电脑里存一首歌那样)，但计算机本身不能自我完成这项工作。而大脑皮层则不同，它可以自我存储这些序列，这是大脑皮层记忆系统的固有特性。

现在，让我们来看看记忆的第二个主要特征——自-联想。正如我们在第二章中读到的，自-联想是指模式可以和自己联系在一起，而自-联想记忆系统是指那种能够根据不完整或混乱的输入信息回忆起全部模式的系统，它对于空间-时间模式都适用。当你看到帷幕后露出的孩子的鞋，就会自然而然地想到整个鞋的样子。是你完善了鞋的部分影像，从而形成了一个完整的空间模式。

另外，想像一个正在等公共汽车的人，因为有灌木丛挡着，你看不到她的全部，但你的大脑并不会因此而混乱。眼睛虽只看到她身体的几个部位，但你的大脑会创造出整个人的感觉，将其他的部位补充完整。这个感觉是如此强烈，你甚至意识不到它是臆想出来的。

你也可以完成时间模式。当你回忆很久以前发生的某个事件的细节时，这个记忆序列就会像潮水一般涌入你的大脑。马塞尔·普鲁斯特(Marcel Proust)的著名系列小说《追忆似水年华》(Remembrance of Things Past)一开始就是回忆美味的玛德琳蛋糕，接着便洋洋洒洒写了 1 000 多页。在一个嘈杂的环境下，我们经常听不清所有的谈话内容，但没关系，我们的大脑

会用它期望听到的东西来弥补那些没有听到的内容；无疑，我们所理解的并不全是我们实际听到的。有些人喜欢大声接别人的话头，而在大脑深处，我们每一个人都在不停地这样做，而且不只是接后半句话，甚至有时是句子的开头和中间。通常，我们意识不到自己正在不断地完善模式，但它确实是无处不在的，它是大脑存储记忆的基本特征。

无论何时，一个记忆片断可以激活全部的记忆，这就是自 - 联想记忆的精髓。

你的大脑皮层就是一个复杂的生物自 - 联想记忆。在你清醒的每时每刻，每一个功能区都始终保持警觉，等待着熟悉的模式或模式片断的输入。即使在你沉思的时候，一看到自己的朋友，你的思绪就会立刻转向她。你不是有意这样去做，但朋友的出现使得你的大脑马上回忆起与她有关的模式。这个反应是自然发生的，无法遏制的。思路被打断后，我们会问：“我刚才在想什么？”在餐桌上和朋友的谈话也会是一种拐弯抹角的联想过程，话题可能会从眼前的食物谈起，而那份色拉可能会让你想起妈妈为你的婚礼做的色拉，从而又想起另一个人的婚礼，由此又想到他们度蜜月的地方，然后又是那个地方的政治问题，等等。

思想和记忆是相互关联的，我又要强调一遍，随机的想法是不可能发生的。输入的信息会自动和自己联系起来，然后填补现有的位置，并自动地和将要发生的事情联系起来。我们将这一连串的记忆称作思想，虽然它的思路仍无法确定，而且我们对它的了解也不完备。

下面，我们来关注一下大脑皮层记忆的第三个主要特性：它是如何形成所谓的恒定表征的。本章节只涉及它的基本概念，我会在第六章中详细地解说它是如何形成的。

计算机存储器存储信息的方法正是它显示信息的方法。如果你将光盘上的程序拷贝到硬盘上，它的每一个字节都和原来的百分之百相同，两个版本中任何一个错误都会使程序失灵。而大脑皮层的记忆则与此不同。大脑记住的并不是我们确切看到、听到或感觉到的；我们之所以不能完全准确地记忆或回忆，并不是因为大脑皮层和它的神经元漏洞百出，而是因为大脑记忆的是世界上各个独立细节之间的重要关系。下面，我要举几个例子来说明这一点。

我曾在第二章谈到几十年前出现的自-联想记忆模型，而且我们的大脑也是利用自-联想方式回想记忆中的东西的，但神经网络研究者们建立的自-联想记忆和大脑皮层的联想记忆之间存在着很大差异。

人造自-联想记忆没有使用恒定表征，因而它在一些最基本的方面无能为力。假设我有一幅画，画上是一张由无数黑白小圆点组成的人脸。这幅画就是一个模式，我可以将许多同样类型的画存储在人造自-联想记忆中。如果我向它提供半个脸或是一双眼睛的画面，自-联想记忆就可以辨认出来，并将缺失的部分正确地补充上去。它很擅长此道，而且这种实验也做过多次。但是，如果我将画面中的每一个小圆点向左移动5个像素，这个记忆系统就无法辨认出这张脸。对于自-联想记忆

来说，由于原来存储的模式和新模式的像素无法吻合，它就变成了一个完全不同的新模式。然而，你我却可以轻易地辨认出这张脸，甚至注意不到其中的变化。如果这些模式被移动或旋转，重新调节了大小或有任何改变，人造自-联想记忆就无法辨认出它们，而我们的大脑却可以轻易处理这些变化。当输入模式发生了变化，我们是如何来感知这一切的呢？让我们来看看另外一个例子。

现在你的手中拿着一本书，当你移动书本，改变照明的灯光，或是调整你的坐姿，或是凝视书的不同版面时，投射在你视网膜上的光的模式就会完全改变。你接收到的视觉输入一刻不停地变化，永不重复。实际上，即使你将这本书在手中捧上100年，视网膜上的模式也不会有任何一次相同，因此，进入你大脑的模式也不相同。即使如此，你也丝毫不会怀疑你正捧着一本书，而且是同一本书。尽管书对大脑的刺激在不断地变化，但你大脑中显示“这本书”的内部模式不会改变。因此，我们用“恒定表征”这个术语来指代大脑内部的表征。

再举一例。想想你朋友的脸，每次看到她，你都能认出她来，这个动作是在不到1秒钟时间内自动发生的，不论她距离你1米还是2米，或者她在房间的另一端，这都没关系。当她离你很近时，她的影像就占据了视网膜的大部分；当她离你很远时，她的影像只会占据视网膜的一小部分。她可以正面对你，也可以略微侧身，还可以侧面对你；她可以微笑，可以眯起眼睛，还可以打着哈欠；她可以在强光下，可以在阴影里，也可以在迪斯科舞厅光怪陆离的灯光下；她的面孔可以出现在不同地点，可以有不同的变化。尽管印在你视网膜上的光

的模式每次都不相同，但你可以立刻感觉到自己注视的是朋友的脸。

现在我们“进入”大脑内部，看看完成这个惊人举动时你的大脑里究竟发生了什么。在实验中监测大脑皮层视觉输入区（即 V1 区）的神经元的活动时，我们发现这些活动模式和她脸的每一个影像都不相同。V1 区中的活动模式随着脸的每一次移动和你眼睛的每一次注视而变化，与视网膜上的模式变化相似。然而，我们在监测你的面孔辨别区域时（面孔辨别区域在脑皮层层级结构中比 V1 区高几个级别），却发现这里细胞活动非常稳定。也就是说，你朋友的脸一出现在你的视线里（即使在想像中），不论它的大小、位置、方向如何，也不论它的比例和表情怎样，面孔辨别区的一些细胞会一直保持稳定的激活状态。这种细胞激活的稳定状态就是一种恒定表征。

从内部看，这个任务似乎简单到不能称其为任务，它和呼吸一样完全是不由自主的，平凡得令人感觉不到它的存在。但从另一个角度看，它之所以平凡，是因为我们的大脑可以很快地解决它（不要忘了“一百步”规则）。然而，对于大脑皮层是如何形成恒定表征的解释，仍然是最大的科学谜团之一。它到底有多难？回答是：即使用世界上功能最强大的计算机，也没人能够解决它。但是，尝试解决这个难题的也不乏其人。

对这个问题的思考由来已久，最早可以追溯到 2 300 年前的柏拉图。他曾提出疑问：人类为什么能够思考，能够了解世界？他指出，真实世界的事物和想法总是不完善、不相同的，比如，你对正圆有一个概念，但在实际中从未见到过正圆，所有画出的圆圈都是不完美的，即使用几何学家的圆规画出的所谓

正圆也有一条黑线，而一个真正圆的周边是没有厚度的。那么你又是如何获得正圆这个概念的呢？或许我们可以换一个更广的角度，想想你对狗的概念。你见过的狗都不尽相同，即使每次看到同一条狗所获得的影像也是不同的。所有的狗都不相同，即使面对同一条狗，你也不可能以完全相同的方式再次看到它。然而，你对狗的不同经验会形成一种稳定的有关狗的思想概念。柏拉图对此大惑不解。

大千世界充满了各种形式和不断变化的感觉，我们是如何从中获得概念的呢？

柏拉图找到了答案，就是他著名的形式论。他认为，我们的高级思维一定被控制在某个超自然的抽象层面上，其中存在着无限完善的稳固思想形式（就是理念，用大写字母 F 代替）。他相信我们在出生前灵魂就来自这个神秘的地方，在那里获得最根本的理念，并且在出生之后保留下来。当真实世界中的各种形式使我们回忆起和它们对应的理念时，就产生了学习和理解。我们之所以有“圆”和“狗”的概念，就是因为它们分别激发了我们大脑中“圆”和“狗”的神秘记忆。

从现代的角度看，这个观点很疯狂。然而，一旦去除掉那些夸张的形而上学的成分，你就会发现他实际上谈论的就是恒定。尽管他的阐述方式与我们的问题毫不相干，但他的直觉却道出了问题的关键，这就是有关我们本质的最重要的问题。

我们再看几个其他有关感觉的例子，以免你产生错觉，认为恒定都是与视觉有关的。首先看看体觉。当你伸手在汽车中

放手套的隔栏里找太阳镜时，只要手指一触碰到它，你就知道自己找到了。是大拇指、手掌或手指的哪个部位接触到的它并不重要；摸到太阳镜的哪个部位也不重要，不论是镜片、镜腿还是镜架。当你手的任何一个部位触摸到太阳镜时，仅仅1秒钟时间就足以让你的大脑能辨别出是太阳镜。在这种情况下，皮肤的部位不同，触摸到的物体部位也不同，因此来自触觉感受器的空间和时间模式流是完全不同的，尽管如此，你却可以不假思索地将太阳镜很快拿出来。

再看一个感觉运动的任务——将钥匙插进汽车点火开关。每一次做动作时，你的座位、身体、手臂以及手的位置都有些许差别，但对你来说，因为大脑中存储着这个动作的恒定表征，它就只是一个日复一日的简单重复动作而已。如果你想制造一个能进入汽车并插入钥匙的机器人，这时你才会明白那几乎是不可能的，除非能确保它每一次的位置和拿钥匙的方式完全一致。即使你做到了这一点，也还要针对不同的汽车给机器人设计程序。而和自-联想记忆一样，机器人和电脑程序在处理变化方面是无能为力的。

另外一个有趣的例子是你的签名。在脑前部的运动皮层有一个亲笔签名恒定表征，每次签名时你使用的有关笔画、角度和节奏的序列都是完全相同的，不论是用流畅的钢笔经常性地签名，还是像约翰·汉考克(John Hancock)那样用胳膊肘在空中炫耀地画一下，还是用脚趾夹着铅笔笨拙地签名，情况是完全一样的。

当然每次都会有少许的不同，尤其是用我刚提过的极端方式。但不论签名的大小如何，书写的工具和身体接触的部位怎

样,你总是使用相同的抽象“运动程序”来完成这个动作。

从签名的例子可以看出,运动脑皮层和感觉脑皮层中的恒定表征从某个层面上看是一模一样的。在感觉方面,各种输入模式可以激活显示某种抽象模式(如朋友的面孔和太阳镜)的稳定细胞群;而在运动方面,显示某种抽象运动命令(如接球、签名)的稳定细胞群又可以遵照各种限制通过大量肌肉群表现出来。如果蒙卡斯尔是对的,大脑皮层在各个区域遵循的是同一个基本规则系统的话,这种感觉和行为之间的对称性就正是我们希望得到的。

最后,让我们以音乐为例,再回到感觉皮层。(我喜欢以音乐为例,因为它可以轻易地帮助我们理解大脑皮层处理的所有问题。)音乐恒定表征是通过你对某个调式的旋律的辨别能力体现出来的。一个曲调弹奏的调式是指旋律的基础音阶。同一首旋律,如果调式不同,开始的音符就不同;一旦你选择了演奏的调式,就确定了旋律中的所有音符。任何一首曲子可以用各种调式演奏,也就是说,用新调式演奏的“同一首”曲子实际上是一系列完全不同的音符。不同版本的曲子对耳蜗的不同部位产生刺激,因而产生出不同的空间-时间模式输入你的听觉大脑皮层,尽管如此,你每次都可以辨别出相同的曲调。除非有非凡的音准能力,否则的话,如果不连续听,你甚至分辨不出这首歌是用不同的调式演奏的。

想想那首《飞越彩虹》。你第一次听到它,可能是电影《绿野仙踪》(The Wizard of Oz)中朱迪·嘉兰(Judy Garland)演唱的,但不可能记得它的调式(降A调),除非你有超人的音准能力。如果我坐在钢琴旁,用你从未听过的调式,如D调,来

演奏，这首歌听起来与原来的没有什么区别。你注意不到所有的音符与你熟悉的那首歌已完全不同了，这就意味着，你对歌曲的记忆忽略了调式。记忆中存储的是歌曲中的重要关系，而不是每一个音符。所谓重要关系是指音符之间的相对音高，即音程。《飞越彩虹》是以一个高八度开始的，接着降了半音，然后又是一个降大三度……这个旋律的音程结构对于任何调式的版本都是相同的。如果你有能力轻易地辨别不同调式的歌曲，说明你的大脑中对这首歌的记忆是以音高恒定形式储存的。

同样，你对朋友面孔的记忆也是以某种独立于特定影像的形式储存的。她的脸之所以被认出，是由脸的相对大小、颜色和比例决定的，与上星期二午餐时某个瞬间的样子没有关系。就像歌曲的音符之间有“音高间隔”一样，脸的特征之间也存在“空间间隔”。对于眼睛来说，脸显得有些宽；和两眼之间的距离相比，鼻子显得有些短；头发和眼睛的颜色之间也有类似的相对关系，即使在不同的灯光下，在绝对颜色变化很大的情况下，这种关系都是保持不变的。你记住了她的脸，就是记住了这些特征。

我相信，大脑皮层的每一个区域都有类似的形式上的抽象，这是大脑皮层的特征。记忆之所以能够被存储是抓住了相互关系的精髓，而不是某个瞬间的细节。当你在看、在感觉、在听某个东西时，大脑皮层就接收到详细的高度特化的输入信息，并将它们转化成一种恒定形式。

实际上，被存储的是恒定形式，与每一个新的输入模式相比较的也是恒定形式。记忆的存储、记忆的唤醒以及记忆的识别都发生在恒定形式的基础之上。在计算机中却没有相

同的概念。

至此，就引出了一个有趣的问题，就是下一章将要谈到的一个大脑皮层的重要功能——如何利用记忆进行预测。既然大脑皮层存储的是恒定模式，它又是如何进行准确预测的呢？下面的几个例子可以解答这个疑问。

假设你生活在 1890 年的美国西部的一个边陲小镇，你的恋人要从东部坐火车来和你结婚，当然你希望去车站接她，于是你提前几个星期就开始关注火车的到站时间。但那时没有固定的列车时刻表，你所知道的只是一天之中火车不会在同一时刻到达或离开，看起来你似乎无法预测她乘坐的火车何时到达。但后来你注意到一个规律，东部来的火车到达时间要比往东去的火车离开的时间晚 4 个小时，即使每天到达和出发的具体时间有很大不同，但中间 4 个小时的间隔是一成不变的。在她到达的那一天，你只要留意东去的列车，看到后就设定闹钟，过 4 个小时到达火车站，她乘坐的列车也刚好到达。这个小故事说明了大脑所面临的问题，也告诉我们大脑是如何解决它的。

你感觉中的世界就像火车到达和离开的时间一样永远不会相同，但是你可以找出存在于不断变化的输入信息中恒定的条理性，并借助它去了解世界。

然而，要进行具体的预测只有这些是不够的，正如只知道火车之间 4 个小时的间隔并不足以使你准时到达站台迎接恋人一样。大脑要准确预测，必须将恒定条理性 and 最近的细节信息

结合在一起。要预测火车到达的时间就必须知道列车时刻的间隔，并将它和细节信息——最后一列东去列车的开车时间相结合。

当你欣赏一首熟悉的钢琴曲时，大脑皮层可以提前判断下一个音符是什么。但正如我们所知，你对这首曲子的记忆是以音高恒定形式储存的，它只能告诉你下一个音程是什么，而不能告诉你下一个音符是什么。

要预测下一个音符，必须将下一个音程和听到的最后一个具体的音符结合起来。如果下一个音程是大三度，你听到的最后一个音符是 C，由此可以推断出下一个音符是 E。你大脑中“听到”的是 E，而不是大三度。除非你辨别错误，或是曲子弹错了，否则你的判断是绝对正确的。

当你看到朋友的脸，大脑皮层会立刻获得信息，并预测有关脸的大量影像细节，要核实她的眼睛准确无误，她的鼻子、嘴唇和头发也和原来的无异。大脑皮层所做的预测极其详尽，甚至可以推测出你都没有见过的方位和环境下有关她脸的低级信息。

如果你对朋友的眼睛和鼻子非常熟悉，你就了解到有关她脸的条理性，并由此可以推断出她嘴唇的位置；如果你知道她的皮肤晒成了棕色，你就可以推测出她头发的颜色。你的大脑正是利用对脸的恒定结构的记忆，并将它与直接经验细节结合起来做到这一点的。

列车时刻的例子正好说明了你大脑皮层中的活动，而音乐和面孔的例子却不是。恒定表征与进行预测的输入信息的结合是确实发生了的，它发生在大脑皮层的每一个区域，无处不



在。是它使你对目前所处的房间做出详细判断；是它使你判断出别人将要说出的话，说话时的语气、口音以及说话声音会来自房间的哪个位置；是它使你准确地知道你的脚什么时间触到地面以及上楼梯时的感觉；是它使你能够接住投来的球，甚至用脚来签名。

本章讨论的有关大脑皮层记忆的三大特征(存储序列、自-联想记忆以及恒定表征)是根据过去记忆推测未来的必要因素，而智能的必要因素——预测，则是第五章将要探讨的主题。



第五章

智能理论新架构

A New Framework of Intelligence



1986年4月的一天，我正在思考“理解”的真正意义。几个月来我为一些基础问题绞尽了脑汁。如果大脑没有触发行为，那它做了什么？当你被动地听演讲时，大脑做了什么？此时此刻，当你阅读这本书时，大脑做了什么？信息进入大脑却不会再出来，到底发生了什么？你现在的所作所为，如呼吸、眼睛的移动等，都是最基本的行为。大脑在清醒时所做的远比你在阅读和理解这些文字时所做的多得多。理解一定是神经活动的结果，那么神经元在理解时又做了什么？

那一天，我环视办公室，看到了椅子、海报、窗户、植物、铅笔以及周围上百种熟悉的东西。扫视时，我的眼睛看到了它们，尽管这种“看”并没有促使我做出任何动作，没有促成任何行为，但我还是“明白了”这个房间以及里面的东西。我正在做塞尔的“中文屋”做不到的事情，不需要从小孔中送出任何东西，我就理解了，但没有任何动作为证。那么，“理解”到底是什么呢？

苦思冥想的我忽然茅塞顿开，那种感觉如此强烈，就像一团迷雾变得豁然开朗。当时我只问了一个问题：如果一个我从未见过的新出现的东西，比如一个蓝色的咖啡杯出现在房间里，将会发生什么？

答案很简单：我会意识到这个东西以前不属于这里，它会让我注意到它是新出现的，我不需要有意地问自己这个咖啡杯是否是才冒出的，很明显它原本就不属于这儿。这个答案看似简单，其中却潜藏着一个非同寻常的构想：要注意到某个事物与众不同，我大脑中一些原本不活跃的神经元开始变得活跃起来。它是怎么知道蓝色咖啡杯是新的，而房间里其他上百件

东西不是呢?这个问题的答案也让我大吃一惊,那就是:我们的大脑在利用记忆不断地对我们看到的、听到的和感觉到的东西进行预测。当我环视房间,我的大脑就利用记忆在我看到某个物体之前进行预测,而绝大部分的预测是在不自觉的状态下进行的,就像大脑的不同部位在说:“电脑摆放在桌子中央吗?是的。它是黑色的吗?是的。台灯在桌子的右上角吗?是的。字典还在我放的位置吗?是的。窗户是长方形的吗?是的。墙与地面垂直吗?是的。阳光是从这个时间的固定方向照进来的吗?是的。”当某种在大脑皮层中没有被储存的视觉模式输入时,这种预测就会被破坏,我的注意力就会被这个差错所吸引。

当然,大脑在预测时并不会自言自语,不会按照某种既定次序进行,也不会对咖啡杯这类确切的物体进行预测。它是用一种平行的方式连续不断地对我们所在环境的构架进行预测。它只会探测那些奇怪的结构,如变形的鼻子或异常的动作。这种下意识的预测无时无刻不在,我们一时无法立刻意识到它,可能这正是其重要性长久以来被我们忽视的原因。它是那么自然而然,轻而易举,以至于我们无法透彻了解大脑中到底发生了什么;它是那么的普遍,因此知觉并不是意识(也就是世界显现在我们眼前的样子)的唯一来源。我们的意识是我们感觉到的一切与源于大脑记忆的预测之结合。

为了说明当时的想法,几分钟后我设计了一个思维实验,叫做“改变了的门”。它是这样的:

每天你回到家时,都会用几秒钟的时间穿过大门。你伸出

手，转动门把手，走进去，然后关上门。这已成为固定的习惯，你经常做，而且很少注意它。假设在你外出时，我溜进你的家，对你家的大门做了一些手脚，比如把门把手向旁边移动几厘米，将球形把手换成指按门闩，或是将黄铜把手换成镀铬把手。我也可以扩大或缩小门框，改变它的颜色，在猫眼的位置安上一个门环，或是加上一个窗户……总之，我可以想出上千种你所不知道的变化。当你回到家，准备开门时，会很快觉察到有些异常。你可能需要几秒钟才能反应上来到底发生了什么，但你却能立刻意识到某种变化。伸出手转动把手时，你会意识到它不在原来的位置；看到门上新开的窗户时，你会觉得怪怪的；如果改变了门的重量，你会觉得使错了劲，感到很惊讶。总之，你会在极短的时间内注意到这上千种变化中的任何一种。

你是怎么做到的？你是怎么注意到这些变化的？要解决这个问题，人工智能和计算机工程师会列出有关门的所有特性并输入数据库，所有门的每一个特性和你家大门的具体特征都设有专门的信息组。当你靠近大门时，计算机就会向数据库询问有关大门的宽度、颜色、大小、门把手的位置、门的重量、声音等信息。从表面来看，这似乎和我环视办公室时大脑核对各种预测的过程很相似，但它们之间的差异是真实而深远的。人工智能的解决方法是不合情理的。首先，提前列出一个门的所有特性是不可能的，那样的话，它将是一份长得没有穷尽的清单；第二，我们还必须将一生中每一秒遇到的每一个物体列出清单；第三，在已掌握的有关大脑和神经元的知识中没有任何一点可以说明神经元是这样工作的；最后，神经元的反应速度

太慢，无法执行计算机类型的数据库，如果是那样，你可能要花费两分钟而不是两秒钟才能注意到门的变化。

因此，只有一种办法可以解释你对改变了的门所做出的反应——你的大脑对于某个特定时刻将要看到、听到和感觉到的东西进行了低级感觉预测，而且是并行的，脑皮层各个区域都会自动地预测下一步的感觉。视觉区对边缘、形状、物体、位置和动作做出预测；听觉区对音调、声音的来源以及声音的模式做出预测；体觉区则对触摸、质地、轮廓以及温度进行预测。

这里所谓的“预测”，就是指参与感觉你家大门的神经元在确实收到感觉输入之前就变得活跃了起来，当感觉输入真正到达之后，再将它和预料的情况进行对比。当你走进大门时，大脑皮层会根据以往的经验形成大量预测；当你伸出手时，它会预测你的手指将会感觉到什么，什么时候你会接触到门；当真正接触到门时，你的关节会是什么角度；当你推开门时，脑皮层会预测门的阻力会有多大，门会发出什么样的响声。当所有的预测都与实际相符合时，你就会走进大门，根本意识不到这些核对预测的过程。对门的预测一旦被破坏，这个差异就会引起你的注意。正确的预测形成了理解——大门与平时无异。而不正确的预测则使人产生疑惑和警觉——门不在原来的位置，门太轻了，门不在中间的位置，门把手的材质不对劲，等等。所有低级预测都是在各个感觉区连续不断地并行发生的。

但这并不是全部。我要提出一个更有力的观点——预测不仅仅是你的大脑所做的事情，它还是大脑皮层的主要功能，也是智能的基础。脑皮层是一个预测器官，如果要解读什么是智能、什么是创造力、大脑是如何工作的以及如何建造智能机

器，我们必须了解这些预测的本质，并搞清楚它们是如何形成的，甚至对作为预测副产品的行为都要透彻地了解。

我不知道是谁第一个提出了预测是了解智能的关键。在科学界和产业界没有人发明出任何全新的东西，人们看到的只是如何将已有的思想纳入新的框架。在被发现之前，新想法的组成成分通常只是在科学对话的背景里传播，而真正的创新是将这些成分组合成一个整体。同样，“脑皮层的主要功能是预测”这一看法也不是全新的，它已存在一段时间了，但在大脑理论和智能的定义方面它却一直没有取得应有的位置。

具有讽刺意味的是，一些人工智能的先驱却试图用计算机建立一个世界模型并进行预测，其中的例子就是 D. M. 麦凯 (D. M. Mackay)。他在 1956 年曾提出，智能机器应该拥有一种“内部反应机制”与接收到的东西相匹配。他没有用“记忆”“预测”这样的名词，但思路是一样的。

20 世纪 90 年代后期开始，像“推论”“生成模式”“预测”等名词悄然进入了科学术语表，它们都是指那些相关的想法。纽约大学医学院的鲁道夫·李纳斯 (Rodolfo Llinas) 在 2001 年出版的新书《旋涡中的我——从神经元到自我》(i of the vortex) 中写道：“对未来事件的结果做出预测的能力对于成功至关重要，极有可能是大脑最根本也是最普通的功能。”另一些科学家，如布朗大学的大卫·曼福特 (David Mumford)、华盛顿大学的罗杰西·拉奥 (Rajesh Rao) 以及波士顿大学的斯蒂芬·格罗斯伯格 (Stephen Grossberg) 都曾著书立说，用各种方式从理论上说明反馈和预测的作用。英国人托马斯·贝叶斯 (Thomas Bayes, 生于 1720 年) 是统计学的鼻祖，以他的名字命名的贝叶斯理论

网络是数学领域的一个分支，专门利用概率理论进行预测。

目前，我们缺乏的是将这些毫不相干的点点滴滴组织起来，形成一个统一的理论框架。我认为，这是我们以前没能够做到的，也是我这本书要达到的目标。

在仔细探讨大脑皮层是如何进行预测的之前，让我们再看几个例子。你会发现，对此思考得越多，你就会越发感到，预测是无处不在的，它是你了解世界的基础。

今天早晨我煎了薄饼。在做饭的过程中，有一次我伸手到台面下打开橱柜的门，不用看，凭直觉就知道我会摸到什么——此时应该是橱柜的把手，也知道什么时候能摸到它。我拧装牛奶的瓶盖，预料到它会转动，然后很轻松地将其打开。我把煎锅放在炉子上，预料到开关会轻轻地按下，然后稍稍费力地转动，1秒钟以后就能听见火苗发出的“噗”的一声。在厨房的每一分钟，我都会做几十个甚至上百个动作，每一个动作又包含许多的预测。我之所以知道这些，是因为一旦这些普普通通的行动中哪怕有一个与预测的动作不一样，我都会很快意识到。

每一次你抬脚走路时，你的大脑就会预测到你的脚何时停止移动以及会给你踩到的东西多少“压力”。上下楼梯时如果一脚踩空，你会马上意识到这个错误；如果你抬脚过低，当你的脚“擦到”台阶的一霎间，你就会知道自己有麻烦了。实际上，你的脚感觉不到什么，是你的大脑在预测，而这些预测和实际的动作没有吻合，才会产生以上的情况。如果是一个电脑

机器人，就会“毫无顾忌地”跌下来，根本意识不到出了什么问题。至于你，你的脚一旦超出大脑判断应该停止的位置哪怕是几厘米，你就会马上意识到。

当你听一首熟悉的旋律时，你的大脑会提前“听”到下一个音符是什么；当你听自己最喜欢的音乐专辑时，每一首歌开始之前你就能知道它的开头是什么。这是怎么回事？原来，大脑中听到下一个音符时应该激活的神经元，在真正听到它之前就提前激活了，于是你在大脑中“听”到了这首歌。神经元的激活是对记忆做出的反应，这种记忆存在的时间长得惊人。多年后当你再次听到这张专辑时，仍然可以在一首歌结束之后自然而然地“听”到下一首歌，这种现象并不少见。但是，如果你不按顺序随机地听你喜欢的一张 CD 中的歌曲，就会有一种因为疑惑而产生的愉悦感，因为你知道自己对下一首歌的预测是错误的。

听演讲时，你通常会提前知道他们要说什么——至少你认为自己知道！甚至有时你听到的不是演讲人实际说的话，而是你期望听到的。（在我小时候类似的事情经常发生，以至于妈妈曾两次带我去看医生，检查我的听力是否出了毛病。）之所以出现这种情况，部分原因是由于人们在谈话时总会使用一些很常见的词语和表达形式。如果听到我说：“白日依山 __，黄河入海 __。”你的大脑神经元就会被激活，在我将每一句说完整之前，就会浮现出“尽”“流”这两个字。（如果你熟悉这首中国唐代诗歌的话，你就会有这种感受。）当然，我们并不总是能够知道别人将要说什么，预测也并不总是准确的，然而，我们的大脑是通过预测可能发生的事情来进行工作的。有时我们可以准确地

知道下一步会发生什么，而有时我们的预测又会产生多种可能性。例如，当我们坐在餐桌旁用餐时，如果我说：“请递给我——”不论下一个词是“盐”“胡椒”还是“芥末”，你的大脑都不会感到惊讶，因为从某个角度来说，它已经预测到所有可能发生的结果了。然而，如果我说：“请递给我人行道。”你马上就会意识到出问题了。

再让我们回到音乐，在这里，我们仍然可以看到对于可能性的预测。当你听一首以前从没听过的歌时，你仍然可以有相当强烈的预测。对于西方音乐来说，我会期望听到某个常规的节拍、某个重复的节奏，或是希望歌曲在主高音上结束。你可能不明白这些术语的意思，但是，假设你在听一种类似的音乐，你的大脑会自然而然地预测节拍、重复的节奏、乐句以及整首歌的结束。如果一首新曲子破坏了这些规律，你会立即意识到曲子有问题。而且当你听一首以前从未听过的曲子时，你的大脑经历的模式是以前从未经历过的，但你仍然可以做出预测，可以辨别出是否出了问题。这些下意识的预测是以存储在你大脑皮层中的记忆为基础的。你的大脑不能确切地说出下一步将会发生什么，但它能够预测出哪一种音符模式可能出现，哪一种不可能出现。

我们都有过这样的经历：当某些连续的背景声音，比如远处的风钻发出持续的噪音，或是在餐馆、商店、工厂等播放绵绵的背景音乐时，我们会毫不在意；然而一旦它们停下来，我们会马上注意到。你的听觉区域预测的是音乐的连续性，只要它不变化，你就不会留意，而一旦它停下来，就破坏了这种预测，于是就吸引了你的注意力。下面是一个真实的故事。纽约

市刚刚停运高架火车时，经常有人半夜报警说他们被什么东西惊醒，而报警的时间往往是以以前火车经过他们公寓的时间。

我们都喜欢说“眼见为实”，而我们看到的事物中有我们真正看到的，也有我们希望看到的。其中一个最有趣的例子和研究人员所说的“填补”有关。你可能早就注意到眼睛的视网膜上有一个小小的盲点，这个盲点上没有光感应器，因而你视觉区域中相应的点也是什么都看不见的。但我们为什么没有注意到它们呢？原因有两个，一个很普通，一个则具有启发性。那个普通的原因，就是两只眼睛的盲点不交叉，因此一只眼睛能补偿另一只眼睛。

但有趣的是，如果闭上一只眼睛，你仍然注意不到盲点，因为你的视觉系统“填补”了丢失的信息。当你用一只眼睛看着华丽的土耳其地毯或樱桃木桌面上波浪形的纹理时，你看不出任何缺陷——毯上的线结和木纹上的黑色结节被盲点覆盖时，视网膜就捕捉不到它们；然而你看到的却是天衣无缝的材质和颜色，因为你的脑视觉皮层利用了相似模式的记忆形成了连续的预测流，填补了丢失的输入信息。

“填补”发生在视觉影像的各个部分，而不仅仅在盲点上。如果我给你看一张海滩的照片，海滩的岩石上放着一些漂流木，木头和岩石之间的界限清晰可见。但是，如果我们放大图片，你就会发现岩石和木头接触点的材质和颜色非常相似。在放大的部分上，木头的边缘很难和岩石分开，但从整个画面看，木头的界限很清楚。实际上这个清楚的界限是我们通过图像的其他部分推想出来的。看世界时，我们可以感到不同物体之间存在的线条和界限，但进入我们眼睛的原始数据经常是

纷乱而模糊的，我们的大脑皮层会将它认为应该的内容填进那些失去或是混乱的部分，这样我们就看到了清晰的影像。

视觉预测也是眼睛移动方式的一种功能。在第三章中我曾提到眼扫视，也就是说每秒钟内有 3 次我们的眼睛会固定在一个点上，然后又突然跳到另一个点。通常你意识不到这种移动，你也不能有意识地控制它们。每一次眼睛变换注视点时，进入大脑的模式都会完全改变，即 1 秒钟内你的大脑会看到 3 次完全不同的东西。眼扫视并不完全是随机的。看到一张脸时，你首先注视的是脸上的一只眼睛，然后是另一只，并不停地打量，然后才会固定在鼻子、嘴、耳朵和其他器官上。你感觉到的是一张脸，而眼睛看到的却是眼睛——另一只眼睛——鼻子——嘴——眼睛，等等。我知道你的真实感觉并非如此，你感觉到的是周围环境的连续景象，但进入你大脑的原始数据却像一台抖动的便携式摄像机所拍下的图像一样不平稳。

想像一下，如果你遇到一个人，他脸上原本是眼睛所在的地方长着一个鼻子。你的眼睛首先会注视一只眼睛，然后是另一只，但如果你看到的不是眼睛而是一个鼻子，这时你肯定知道出了问题。之所以会这样，是因为你的大脑对于将要看到的東西有预测。当你预测的是眼睛，而看到的却是鼻子，这个预测就被破坏了。因此随着每一次眼扫视，你的大脑在每秒钟都会对下一步将要发生的事情进行多次预测，如果哪个预测出现了错误，马上就会引起你的注意。这也解释了为什么遇见身体畸形的人我们总是忍不住多看几眼。如果看到一个长着两个鼻子的人，你很难做到眼睛不盯着看。当然，如果和这个人一起生活一段时间后，你就习惯了这两个鼻子，也就不会再把它当

作什么不寻常的事情加以关注了。

现在，想想此时此刻的你在做什么预测。当你翻开书本，你会预测到页码会略微弯曲，并以某种可预测的方式翻动，这种方式与封面翻动的方式完全不同。如果你坐在椅子上，你会预测到身体感觉到的压力是持续不断的。但如果椅子突然湿了，或者开始向后滑动，或者是什么突然的变化，你的注意力就会从书本上移开，尽量要弄明白发生了什么事。如果你花些时间观察一下自己，就能明白你对世界的看法、对世界的理解同样也和预测紧紧联系在一起。你的大脑早已建立起了一个有关外部世界的模型，并不断地将这个模型和事实相比较，因为这个模型的有效性，你才会知道自己在哪里，自己正在做什么。

预测并不仅仅存在于“看和听”这一类低级的感觉信息模式上。到目前为止，我的讨论一直局限于这类事例，是因为这样的例子是介绍智能新框架的便捷方式。然而，根据蒙卡斯尔的理论，低级感觉区感知到的事实，对于脑皮层的所有区域来说，都是普遍存在的。人类的大脑之所以比其他动物的大脑聪明，是因为它能够对更加抽象的模式和更长的时间模式流做出判断。现在，我的妻子正盯着我，要想预测她要说什么，我必须知道她说过什么——今天是星期五；垃圾袋必须在周五的晚上放到马路边上；上个星期五我把这事忘了。现在她脸上是那样一种表情。当她张开嘴时，我有强烈的预感，知道她要说什么。在这种情况下，尽管不清楚她的具体措辞，但我确实知道她要提醒我把垃圾袋拿出去。我要说的是，高级智能并不是一种不同于感性智能的处理过程，它同样是以大脑皮层记忆和预测算法为基础的。

要知道，智力测试从本质上讲就是预测能力的测试。从幼儿园的游戏到门扎(Menza)测验，所有智商测试的基础都是做预测。给你一个序列数，下一个数字应该是什么？给你看一个复杂形状的物体的三个面，下面哪一个也是这个物体的面？以词语 A 和词语 B 的关系为前提，那么词语 C 和哪个词有类似的关系？

科学本身就是一种预测练习。我们进行一系列的假想和实验，从而不断改进我们对世界的看法。这本书也是对什么是智能以及大脑工作原理的预测。甚至连产品的设计也是一种基本的预测过程。不论是设计服装还是设计手机，工程师们都会猜测竞争对手的做法、消费者的需求、新产品的成本以及什么样的款式是市场需要的。

智能是通过记忆能力和对周围环境中模式的预测能力来衡量的，包括语言、数学、物体的物理特性以及社会环境。你的大脑是从外部世界获得信息并将它们储存起来，然后将它们以前的样子和正在发生的情况进行比较，并以此为基础进行预测的。

说到这儿，你可能会想：“即使躺在黑暗中什么都不做，我也是智能的。这个说法我可以接受。正如你所说的，我并不需要专门做什么来理解或获得智能。但这样的情况难道不是例外吗？你难道能说智能的理解与行为毫不相干吗？最终使我们具有智能的不正是行为本身而非预测吗？毕竟，行为是生存的最终决定者。”

这个问题提得很好。当然，对于动物的生存来说，行为是最重要的因素。预测和行为不是完全分离的，但它们之间的关

系相当微妙。首先，在进化过程中，新大脑皮层是在动物具有复杂行为之后才出现的，因此理解脑皮层存在价值的基础是动物现有行为的不断改进。也就是说，先有行为，然后才有智能。第二，除了听，我们对世界的感知在很大程度上依赖于我们的运动方式，因此，预测和行为又有着紧密的联系。让我们看看下面的例子：

哺乳动物通过进化得到一个很大的新大脑皮层，并带来某种生存优势，而这种优势最终会根植于行为之中。然而在最初，脑皮层的作用只是如何更好地利用已经具有的行为，而不是创造出全新的行为。为了更清楚地解释这一点，我们需要回顾一下人类大脑的进化历程。

几亿年前，多细胞生物开始在地球的各个角落繁衍，不久以后，就出现了简单的神经系统。然而真正的智能是随着我们的爬行类祖先的出现开始的。随后，爬行动物很成功地征服了整个大陆，它的踪迹遍布各个大洲并分化出无数的物种。它们有着很敏锐的感觉和能够赋予它们复杂行为的、相当发达的大脑。它们的后裔——现在的爬行动物，也继承了它们老祖宗的这些能力，就像短吻鳄，它拥有和你我一样的复杂感觉，有发达的眼睛、耳朵、鼻子、嘴和皮肤，也具有很复杂的行为能力，如游泳、奔跑、隐藏、捕猎、伏击、晒太阳、筑巢和交配。

那么，人脑和爬行动物的大脑到底有什么差别？可以说它们的差别既显著也微小。说它微小，是因为粗略地看，爬行动物大脑中的东西人脑中都有。说它显著，是因为人脑中有某些爬行动物没有的真正重要的东西，那就是一个很大的新大脑皮

层。有时候你会听到“古脑”或“原始脑”这类的说法，像爬行动物一样，每个人的大脑中都有这些较古老的结构，它们控制着人的血压、饥饿感、性欲、情感以及运动的各个方面。例如，当你站立、行走或尽力保持身体平衡时，你主要依靠的是古脑；如果你被声音吓了一跳，感到惊慌并开始奔跑，那也是你的古脑在起作用。要做这一类有趣而有用的事情，一个爬行动物的大脑就足以胜任。那么，如果不严格地要求它去看、去听或去运动，新大脑皮层又会做些什么？

哺乳动物有新大脑皮层，因此它们比爬行动物更聪明。（“新大脑皮层”源于拉丁语，意思是“新长出的树皮”或“新的外层硬皮”，因为它的确将古脑包裹了起来。）几千万年前，新大脑皮层首先出现在哺乳动物身上，几百万年前才得以显著发展，正是这一阶段的发展使得我们人类比其他哺乳动物更聪明。不要忘了，新大脑皮层的形成使用的是很普通的重复性元素，人类大脑皮层的厚度和结构都与我们的哺乳动物亲戚差不多。当大自然要快速地进化某个重要的东西时，例如人类大脑皮层，它就会复制某种已经存在的结构，并加入更多的有关普通脑皮层算法的元素，我们才得以获得聪明才智。人们有一种普遍的错误观念，认为人类大脑是几十亿年进化的顶峰。对于整个神经系统来说的确如此，然而，新大脑皮层本身也是一个相对较新的结构，并没有经历长时期的进化过程。

如何理解新大脑皮层？为什么记忆和预测是打开智能谜团的钥匙？下面就要谈到这些问题的核心。让我们从没有脑皮层的爬行动物的大脑开始吧。进化论发现，给原始脑的感觉途径增加一个记忆系统（也就是大脑皮层），动物就能够获得预测未来的

能力。这样的话，让我们想像一下，当爬行动物的古脑在执行它原有功能的同时，现在的感觉模式就会自动传输到新大脑皮层，并在记忆系统中储存起来。将来的某个时间，当这个动物遇到相同或相似的情况后，记忆系统就会辨认出相似的输入并回忆起以前发生的事情。重新唤起的记忆和感觉输入信息流对比之后，大脑就可以“填补”目前的输入信息并对下一步将要看到的情况进行预测。通过对现实的感觉输入和唤起的记忆的对比，动物不但可以了解自己所在的位置，并且可以预见到未来。

假设大脑皮层不仅能记住已经看到的東西，而且能记忆处在相似情况下古脑所做出的行为，我们甚至不必去假设大脑皮层能够区别感知和行为，因为对大脑皮层来说，它们都是模式。当动物发现自己处在相同或相似的环境中时，它不但可以预测未来，而且能够回忆起什么样的行为可以产生什么样的结果。这样，记忆和预测使得动物可以更聪明地利用已经存在的（古脑的）行为。

举个例子。假设你是一只老鼠，第一次学走迷宫，因为你不知所措又饥饿难忍，你一定会利用古脑固有的技能在这个新环境中摸索——去听、去看、去嗅并顺着墙壁向前爬。所有这些感觉信息都是由古脑所控制，然后输入到你的新大脑皮层并储存起来。之后的某一天，当你身处相同的迷宫时，你的大脑皮层会辨别出当前的输入信息与以前看到的相同，于是就回忆起已存储的代表着过去发生情况的模式。从本质上看，正是它使你能够看到不远的未来。如果你是只会说话的老鼠，你一定会说：“哇，我认出了这个迷宫！我记得这个拐角！”当你的大脑皮层回忆起以前的情况时，你的眼前就浮现出上次在迷宫中

看到奶酪并向它奔去的情景。“从这儿向右拐，我知道下面会发生什么。这条通道的尽头有一块奶酪，我已经能想像出它的样子了。”当你匆忙穿行于迷宫时，你所做的动作，诸如抬腿、理胡须等，都是依靠你的古脑皮层的原始结构来完成的。然而，你之所以能够记得曾到过的地方，并在将来的某个时候再次认出它，而且对将要发生的事情进行预测，这些都要归功于你的(相对较大的)新大脑皮层。蜥蜴没有大脑皮层，因此它对过去的记忆能力就差得多，每一次走迷宫时它都会一遍又一遍地搜寻。你(作为一只老鼠)之所以能了解世界并预知不远的将来，都是因为你拥有大脑皮层记忆系统。你能生动地回忆起每一次对于摆在前面的奖励和危险的判定，因而可以更为有效地行动。也就是说，你确实可以预见到未来。

但是请注意，你并没有做出任何特别复杂或从根本上看前所未有的全新行为，你并没有把自己变成一架滑翔机飞向通道尽头的奶酪。你的大脑皮层形成的感觉模式预测使你能够预见到未来，但你所有的一切行为并没有受到影响，你所有的能力，像乱窜、攀爬和摸索等，和一只蜥蜴没有差别。

大脑皮层进化得越来越大，能够记忆越来越多的东西，能够形成更多的记忆，进行更多的预测，而且这些记忆和预测的复杂性也越来越强。除此之外，在人类脑皮层上还发生了更加不可思议的变化，形成了人类所特有的智能行为能力。

人类的行为远远超出了诸如老鼠的技能之类原有的基本功能。人类大脑皮层的进化已达到了一个新的高度，因为只有人类创造了书写和口头的语言，只有人类能够蒸煮食物、制作服装、开飞机、建造摩天大楼。我们的运动和谋划能力也大大超

过了与我们最亲近的动物亲戚。天生能进行感觉预测的脑皮层是如何创造出人类独有的极其复杂的行为呢?这种超级行为何以会如此突然地得到进化呢?对此我有两种答案:一是因为脑皮层算法极其强大灵活,只要线路进行一点点专属人类的变化,就能创造出全新的复杂行为。另一种回答是,行为和预测实际上是同一事物的两个方面,尽管大脑皮层能够预测未来,但它只有在对所发生的行为有所了解的条件下,才能做出准确的感觉预测。

在老鼠找奶酪的例子中,老鼠记得那个迷宫并且可以利用这个记忆预测到自己会在拐角处看到奶酪。它可以向左转也可以向右转,但只有同时记起奶酪和它所做的正确行为——在岔路口向右拐,才能使吃到奶酪的判断变成现实。虽然这个例子很简单,但它触及了感觉预测和行为之间紧密关系的精髓——所有的行为都能改变我们所看到、听到和感觉到的,而我们随时随地得到的大部分感觉又是由行为决定的。在你的面前动一动手臂,你的大脑皮层必须知道它发出了移动手臂的命令才能做出“看见手臂”的预测。如果大脑皮层没有发出相应的运动命令你却看见手臂在动,你一定会惊讶不已的。对此最简单的一种假设方法是:你的大脑首先命令手臂移动,然后才能预测到将要看到的景象。我认为这种解释是错误的。我的看法是脑皮层能预测到看见了胳膊,是这种预测引发了运动命令并使预测变为现实,也就是说,你首先想,然后促使自己去行动来实现这个想法。

现在,我们来看一看是什么样的变化使人类的能力得到如此大的发展。猴脑皮层和人脑皮层之间的物理差别可以解释人类拥有语言和其他复杂行为的原因吗?人类大脑的体积是大猩猩

的3倍，但“越大越好”是远远不够的。人类行为的发展之所以有如此大的跨越，其答案在于脑皮层各个区域和古脑各个部分之间的线路连接，简单地说就是我们大脑的连接有所不同。

让我们再仔细地研究一下。大脑的左、右两个半球是每个人都熟悉的，但还有一个部位并不为人熟知，而这个部位正是人脑独特性之所在。所有的大脑，尤其是体积较大的，都将脑皮层分为前、后两个部分，中间有一条很大的裂纹，称作“中央回间沟”。科学家用“前部(anterior)”和“后部(posterior)”加以区别。从眼睛、耳朵和触觉传来的输入信息都进入脑皮层后部，这里是大部分感觉预测发生的地方。脑皮层前部的区域主要涉及高级策划和思维，也包括运动皮层以及负责运动肌肉和做出行为的大脑部位。

灵长目动物的脑皮层已进化得越来越大，其前部的发展更是超出了比例，尤其是人类脑皮层。和其他灵长类动物以及早期类人猿相比，我们人类有很大的额头以承载很大的前部脑皮层。但这种变化并不足以解释和其他生物相比我们在运动能力上的优势。人类之所以拥有可预测的复杂运动能力，原因在于我们的运动脑皮层和我们身体上的肌肉有更多的连接。在其他哺乳动物身上，脑皮层前部在运动行为方面发挥的直接作用并不是很大，它们主要依靠古脑的各个部位产生行为。相比之下，人类脑皮层在运动控制方面取代了大脑的其他部分。如果一只老鼠的运动皮层受到损伤，可能不会有明显的运动能力的损失；而一个人的运动皮层受到损伤后，他就会瘫痪。

人们经常问我，难道海豚有体积较大的大脑吗？答案是肯定的，而且海豚有一个很大的新大脑皮层。但和人类的相比，它

们的脑皮层结构较简单(只有3层,人类的有6层),但用其他标准来衡量,它们的大脑是很大的。海豚可以记住和理解很多事情,可以认出自己的同类;它们对自己的生活有很好的记忆,带有一种自传的意味,它们甚至还记得曾经到过的海洋的每一个角落和裂缝。然而,尽管它们有一些很复杂的行为,但无法和我们人类相比。由此我们可以推测它们的脑皮层对它们的行为并没有决定性的影响,也就是说,它们的脑皮层的主要功能是提供对周围世界的记忆。如果一个动物拥有很大的脑皮层,它就可以和你我一样感知这个世界,但人类脑皮层在控制行为方面所起的作用是决定性的、极其超前的,这是人类所特有的,是我们能够使用复杂语言和制造精密工具的原因,也解释了我们为什么能够写小说、上网冲浪,为什么能向火星发射探测器,为什么能探索神秘能源的性质。

现在一幅完整的图画展现在我们眼前。大自然首先创造出一些动物,它们有复杂的感觉,相比之下其行为却相对僵化,诸如爬行动物之类。然后,只要给它们加入一个记忆系统并输入感觉信息流,这些动物就可以记住以前发生的事。当它们身处相同或相似的环境时,这些记忆就可以被唤起,引发对下一步的预测。这样,智能和理解首先从记忆系统开始,并将预测输入感觉信息流。这些预测就是理解的实质,因为知道某个事情就意味着你可以对它做出预测。

大脑皮层是从两个方面进化的:一是,在储存记忆的形式上向更大、更复杂的方向发展,使得它可以记住更多的东西,可以根据更复杂的关系做出判断。二是,脑皮层首先与古脑的运动系统相互作用。要预测将要看到、听到和感觉到的,大脑

皮层必须要知道将采取的动作是什么。脑皮层控制了人类大多数运动行为，它不仅仅根据古脑的行为做出预测，而且还引导行为去迎合它的预测。

人类脑皮层很大，因而有相当巨大的记忆容量，它能够不断地预测你将要看到、听到和感觉到的东西，而且多数都是在你不自觉的情况下发生的。这些预测就是我们的思想，与感觉输入信息流结合之后就形成了我们的知觉。我认为对于大脑的这个认识，就是智能的记忆和预测框架。

如果塞尔的“中文屋”有一个相似的记忆系统，可以预测到下一个汉字是什么，故事下一步会发生什么，我们就可以肯定地说，这个房间懂汉语，可以理解这个故事。现在我们可以清楚地看到阿兰·图灵错在哪里——智能的证据是预测，而不是行为。

至此，我们已经做好了准备，进一步探讨这个记忆和预测框架的新想法。要对未来发生的事做出预测，你的大脑皮层必须储存一系列的模式；要唤醒相应的记忆，必须根据新旧模式之间的相似性检索这些模式（自-联想回忆）；最后，这些记忆还必须以恒定的形式储存起来，这样对于过去事件的认识才能应用到与之相似但不完全相同的情景中去。我们的大脑是如何完成这些任务的？在下一章，我将回答这个问题，并对大脑皮层的层级结构进行全面的探讨。

第六章

大脑皮层工作原理

How the Cortex Works



要弄清楚大脑是如何工作的，就好像是在玩一个巨大的拼板游戏。你可以采用以下任何一种方式来解决这个问题。一是使用“自上而下”的方法，即从最终拼好的形状着手，开始在那一堆拼图中决定如何取舍。另一种方法是“自下而上”，也就是说，集中关注每一块拼图自身，研究它们各自的特点，以寻求它们与其他拼图块的最佳匹配。如果你不知道拼出的图形是什么样子，那么“自下而上”法就是解决问题的唯一途径了。

像“了解、掌握大脑”这样的拼图游戏是尤其令人沮丧的。因为缺乏一个理解人类智能构成的成熟的理论框架，科学家就只能采用“自下而上”的方法了。要解答“大脑是如何工作的”这样复杂的难题，如果说不是不可能的话，那工作量也是相当大的。设想一下一个由几千块拼板组成的拼图游戏，便可知其困难的程度了。其中许许多多的拼块有多种拼法，就好像是每一块拼图两面都有图案，而只有一面是正确的，而且所有的拼图形状都不太规则，你甚至无法确定两块之间能否进行拼接。还有许多拼图最终都是多余下来的，而你却不知道哪些用不着、到底有多少是会剩下来。制造拼图的厂家每个月都寄来一些新的拼块。这些新的拼块中有一些是用来替换旧的，就好像拼图制造商这样说：“我们知道你用这些旧的拼块拼了多年，但实际上它们是被弄错了。对不住。在没有接到进一步的通知之前，请用这些新的拼块。”不幸的是，你都不知道最后会拼出个什么东西；还有更糟糕的，你也许有了一些想法，但这些想法都是错的。

这个拼图的类比很好地描述了我们在建构关于大脑皮层和智能的新理论时所面临的困难。这些一块块的拼图便是科学家

一百多年来收集的生物学和行为学研究的数据。每个月都有新的论文发表，从而又添加了新的拼块。有时某个科学家的研究数据会与另一个科学家的相冲突。因为可以用不同的方式对所获得的数据进行解释，所以科学家几乎凡事都存在分歧。没有一个“自上而下”的理论框架，对于要探寻什么、什么是首要的、怎样解释堆积如山且越来越多的信息等问题，就没有一个一致的意见。对大脑的深入理解利用“自下而上”的方法是行不通的。我们所需要的就是一个“自上而下”的理论框架。

“记忆 - 预测模型”可以扮演这个角色。它可以使我们知道如何着手将拼块拼合到一起。为了进行预测，大脑皮层必须能够记住并且存储关于事件序列的知识。要预测新的事件，大脑皮层必须形成恒定表征。大脑需要按照世界的本来面貌建立并存储一个世界的模型，这个模型与你在变化着的条件下看到它的样子无关。了解了大脑皮层该做些什么，可以引导我们理解它的结构，尤其是它的层级结构和分为 6 层的形状。

在此首次提出的新框架的时候，我们将就一些对某些读者来说具有挑战性的细节展开详细地讨论。读者将遇到许多陌生的概念，这些概念甚至对神经科学家来说都不太熟悉。但我相信，经过一些努力，任何人都可以领会这一新框架的基本要义。本书第七章、第八章的内容不会太专业，在那两章中我们将讨论本理论的更深广的蕴意。

我们玩拼图游戏的过程现在已经变成了寻找支持记忆 - 预测假设的生物学细节，这就像是我们可以把大多数拼块放到一边，因为我们知道剩余的为数不多的拼块将揭示最终的答案。一旦我们知道了要找寻什么，这一工作就具有可操作性了。

同时我想强调的一点是，这一新的理论框架仍然不是完整的，其中还有许多东西我自己也不甚了解。但是我对理论框架中的许多内容却已掌握，这些内容都是以演绎推理、许多在不同的实验室里进行过的实验以及已知的解剖学知识为基础的。在过去的 5~10 年间，神经科学内部许多子领域的研究人员也在探究某些观念，正好与我所提出的观念相类似，而且，虽然他们用的术语不一样，但据我所知，他们并没有尝试着把这些思想整合到一个完整的框架里。他们也讨论“自上而下”和“自下而上”的处理方式，也讨论模式是如何经由大脑的感觉区域进行传递的，他们还讨论恒定表征可能会是如何的重要。比如说，加州理工学院的两位神经科学家加布里埃尔·克里曼 (Gabriel Kreiman) 和克里斯托弗·科赫 (Christof Koch) 与神经外科医生一起发现，一个人不论何时看到比尔·克林顿的图像，他的某些脑细胞都会变得活跃起来。我的目标之一就是解释这些有关克林顿的细胞是如何形成的。当然，一切理论做出的预测都要能在实验室得到检验。我在附录中列举了许多这样的预测。既然我们已经知道了要寻找的目标，这个相当复杂的系统看上去也就不再那样复杂了。

在本章随后的几个部分里，我们将一步步深入地探讨大脑皮层的记忆 - 预测模型是如何工作的。我们从新大脑皮层的主要结构和主要功能开始，直到最终弄清楚那些细小的拼块的作用以及它们是如何置入整个大画面的。

恒定表征

在前面，我把脑皮层比作是一块餐巾大小、有 6 张名片厚

度的由细胞组成的薄片，在这张薄片上，各个区域之间的连接构成了一个层级结构。现在我想为大脑皮层描绘出另一幅图像，以此来突显它的层次连通性。让我们想像一下，我们根据不同的功能将“餐巾”切成彼此独立的区域(大脑皮层的这些区域都专门负责某类任务)，然后将这些分块像薄饼一样一层层叠起来。要提醒你的是，脑皮层看起来其实并不是这样，但这幅图像能帮你想像信息是如何传递的。我在图上标示出了4个脑皮层区，其中感觉输入经由底层，也就是最低的那个区进入，然后向上逐层传递。请注意，信息是双向传递的。

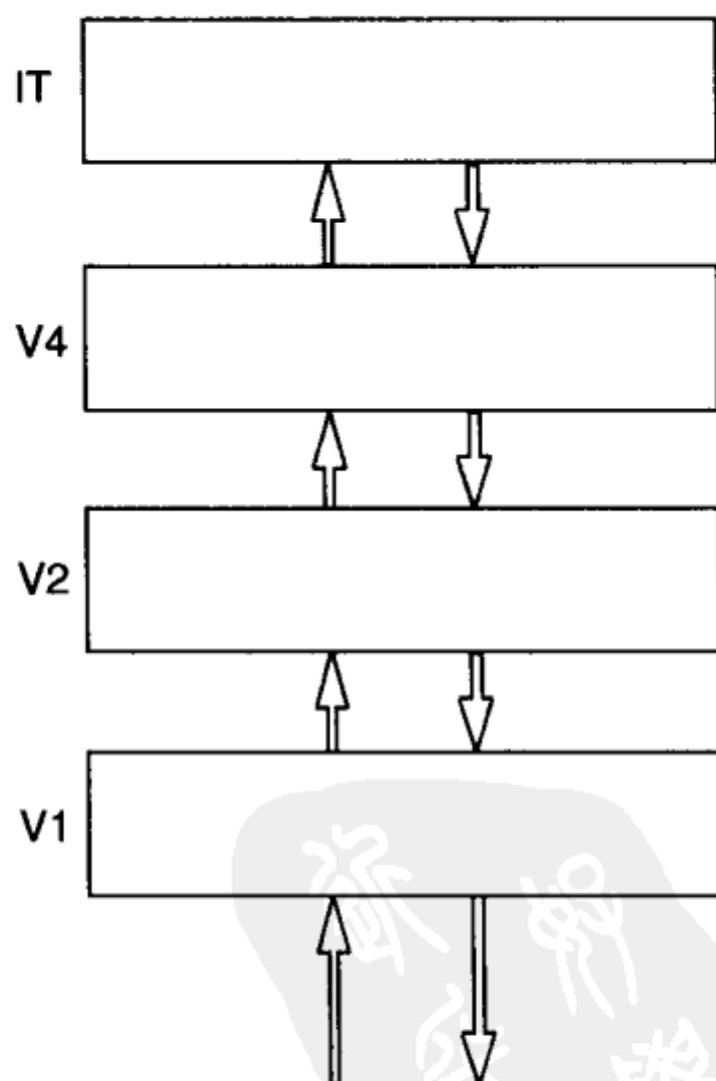


图1 识别物体时前四个视觉区域

图1表示了识别物体时涉及的前4个视觉区域——你是如何看到并辨识出一只猫、一座大教堂、你的母亲、中国的万里长城以及其他任何一个你能叫得上名字的事物的。生物学家把

它们标记为 V1 区、V2 区、V4 区和 IT 区。在图 1 的底部，我们用箭头表示了源于双眼视网膜的视觉输入，这些输入被传递到 V1 区。我们可以把这些输入看作是由约 100 万个轴突（这些轴突束成视神经）所承载的不断变化着的模式。

前面我们讨论过空间及时间模式，在此还是有必要请你重新回忆一下，因为我们将要反复地提到它们。让我们来回忆一下，脑皮层是一大块细胞组织，其中包含有各司其职的功能区域，这些区域由大束轴突或纤维连接起来。信息就沿着这些轴突或纤维瞬时从一个区域传递到另一个区域。在任何一个时间点上，一些纤维会触发出叫作“动作电位”的电脉冲，而其他的纤维则保持静止不动。这样一束纤维上的集体活动就是我们所说的“模式”。

如前所述，人的眼睛每秒大约以 3 次扫视的视觉运动移过某一物体，并有 1 次注视的视觉停留。假若科学家给你安上一个跟踪记录眼球运动的装置，你会吃惊地发现扫视是多么地不平稳。因为之前你体验到的视觉是连续且稳定的。

图 2(a)显示了人眼在看一张脸时是如何运动的。请注意：眼球在注视物体时并不是随机的。现在请想像一下，你能够看见从此人眼中发出并到达 V1 区的活动模式。眼球的每一次扫视都会完全改变。视觉皮层以每秒数次的频率看到一个崭新的模式。

你也许会想：“不错。但眼球所看到的仍然是同一张脸呀，只不过是有所移位而已。”这样想也是有些道理的，但实际情况不全是你所想的那样。视网膜上的光感应体并不是均匀分布的，中心部分分布密集，叫中央凹，四周逐渐稀疏。与此相

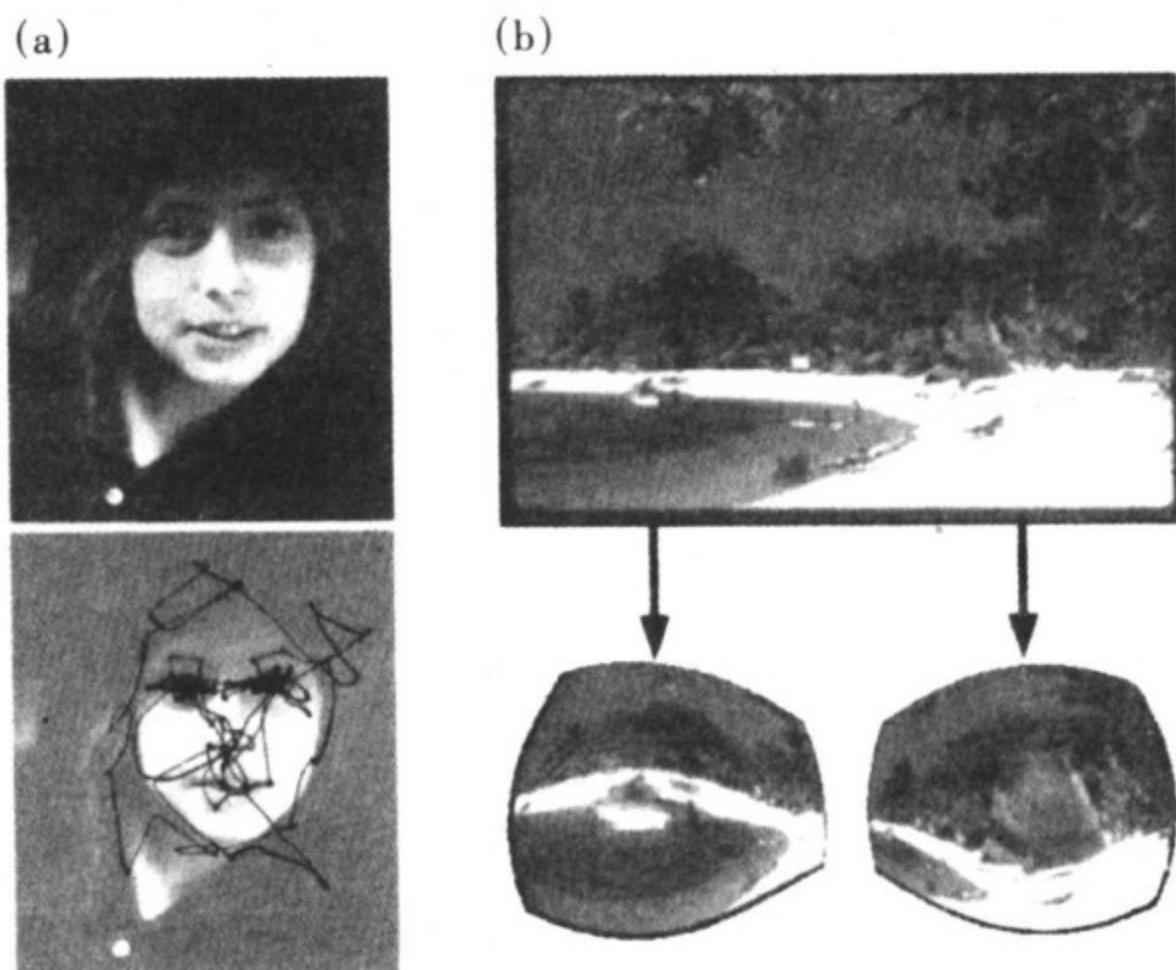


图2 (a)眼睛是如何扫视一张人脸的；(b)由视网膜上感应器分布不均引起的变形

反，脑皮层中的细胞则是均匀分布的。结果，传递到第一视区V1区的视网膜上的影像就被高度扭曲了。当你的眼睛注视某张脸上的鼻子而不是眼睛时，视觉输入是相当不同的，就好像是在前后急拉一个扭曲变形的超广角镜头并通过它进行观察一样。然而当你看那张脸时，它看上去并没有变形，也不跳来跳去。大多数时候，你根本就不知道视网膜的模式有所变化，更不用说那么显著的变化了。你看到的只不过是张“脸”。图2(b)显示了一幅此效果下的海滩风景。在此，我们只不过是再次说明了在第四章关于记忆时我们曾谈到的恒定表征的奥妙。你感觉到的并不是V1区看到的。你的大脑究竟是如何知道正在看的是同一张脸，而你又为何不知道输入有所变化而且变形了呢？

如果将一根探针深入到V1区来看看单个细胞是如何响应

的，我们就会发现任一细胞都只针对视网膜一极小部分的视觉输入做出反应，且活跃异常。这样的实验进行过多次，是视觉研究的主要成果。每个 V1 区神经元都有一个感应域，它只针对你整个视觉区域的很小一部分起作用。说得更精确些，这个区域也就是你的视野范围。V1 区的细胞似乎就不知道什么脸啊、汽车啊、书啊以及你时刻看到的任何其他有意义的物体，它们所“知道”的一切仅仅是视觉世界的一个针孔大小的部分。

V1 区的每一个细胞也都只响应特定类型的输入模式。例如，在某个细胞的感应域范围内，当它“看”到倾斜成 30 度的一条直线或边缘时，这个细胞就会异常活跃。这条直线或边本身并没有什么意义，它可以是任何事物的一部分——地板、远处的棕榈树树干、字母 M 的边线，或任何能想到的其他可能的事物。随着每一次新的“注视”，细胞的感应域就会落在视觉空间的一个全然不同的、新的部分。在某些“注视”中，这个细胞就会活跃，而在另外一些“注视”中，它就不怎么活跃，或者甚至没有反应。因此，眼睛每扫视一次，V1 区的许多细胞就可能要改变它们的活动。

然而，如果我们将探针深入图 1 中的顶端区域，也就是 IT 区的话，奇妙的事情就会发生。我们发现，当整个物体在视野的任何地方出现时，某些细胞会变得活跃起来，并一直保持活跃。例如，我们会发现，每当看到一张脸时，有一个细胞总会被激烈地触发，只要你的眼睛一直看着视野中任何一处的一张脸，这个细胞就会一直保持活跃，它不会像 V1 区细胞那样会随着每一次扫视而不断切换状态。这个 IT 区细胞的感应域覆盖

了整个视觉空间，并且每当看到脸时就会被触发。

让我们再来看看这神秘之处。从视网膜到 IT 区在跨过脑皮层的 4 个阶段的过程中，细胞在改变着状态：从飞速变化的、具有空间特异性、能识别细微特征的细胞，转变成为不断被触发、空间非特异性、能识别目标的细胞。IT 区的这个细胞告诉我们，在视野中我们看到了一张脸。这个细胞通常被叫做“脸细胞”。不管看到的脸是倾斜的、旋转的还是部分被遮挡了的，它都会活跃。它构成了“脸”的恒定表征的一部分。

如此描述使得它看似非常简单。也就是说，经历这么 4 个飞快的阶段，我们就识别了一张脸。任何计算机程序和数学公式都丝毫没能像人脑那样强有力地、高度概括地解决这个问题，而人脑只需几步就解决了这个问题。本章的主要目的之一就是解释这样一个“脸细胞”——不管是关于克林顿还是其他任何人的——是如何出现的。

我们再来看一下图 1，其中可以看到信息通过一个“反馈连接”的网络同时从较高的区域传递到较低的区域。这个网络是由许多束轴突组成的，它们连接着 IT 区这样的较高区域和 V4 区、V2 区、V1 区这样的较低区域。此外，在视觉皮层中，如果说这样的反馈连接不比前馈连接多的话，至少也是一样多。

许多年来，大多数科学家都忽视了这些反馈连接。如果我们对大脑的了解仅仅集中在脑皮层是如何接受所输入的信息、如何加工信息，然后如何对它做出反应的话，就不需要反馈，而仅仅是连接着感觉器官和脑皮层运动区的前馈连接。但是当我们开始认识到脑皮层的核心功能是做出预测时，就会不得不将反馈纳入我们的模型里来了。大脑要将信息送回最先接受输入

的区域。预测就是在实际发生了什么和我们期待发生什么之间进行比较。信息向上流动，我们的期待向下流动。

在涉及所有感觉的皮层各区都存在着同样的前馈、反馈过程。图3所示为视觉“薄饼”与听觉和触觉“薄饼”并排堆叠。图中还画有几个较高层次的脑皮层区，这些叫做联合区的皮层区，接收诸如听觉、触觉和视觉等的输入，并将这些输入进行整合。图1是根据4个已知的脑皮层区域之间已知的连接而绘出的。但图3仅仅是一个示意图，并不是要描绘实际的皮层区域。在真正的人脑里，数十个皮层区域都是通过各种方式互相连接起来的。实际上，人类脑皮层的大部分是由联合区组成。这里的图示以及以后所见到的图解都是为了帮助读者理解，我不希望给读者造成任何的误解。

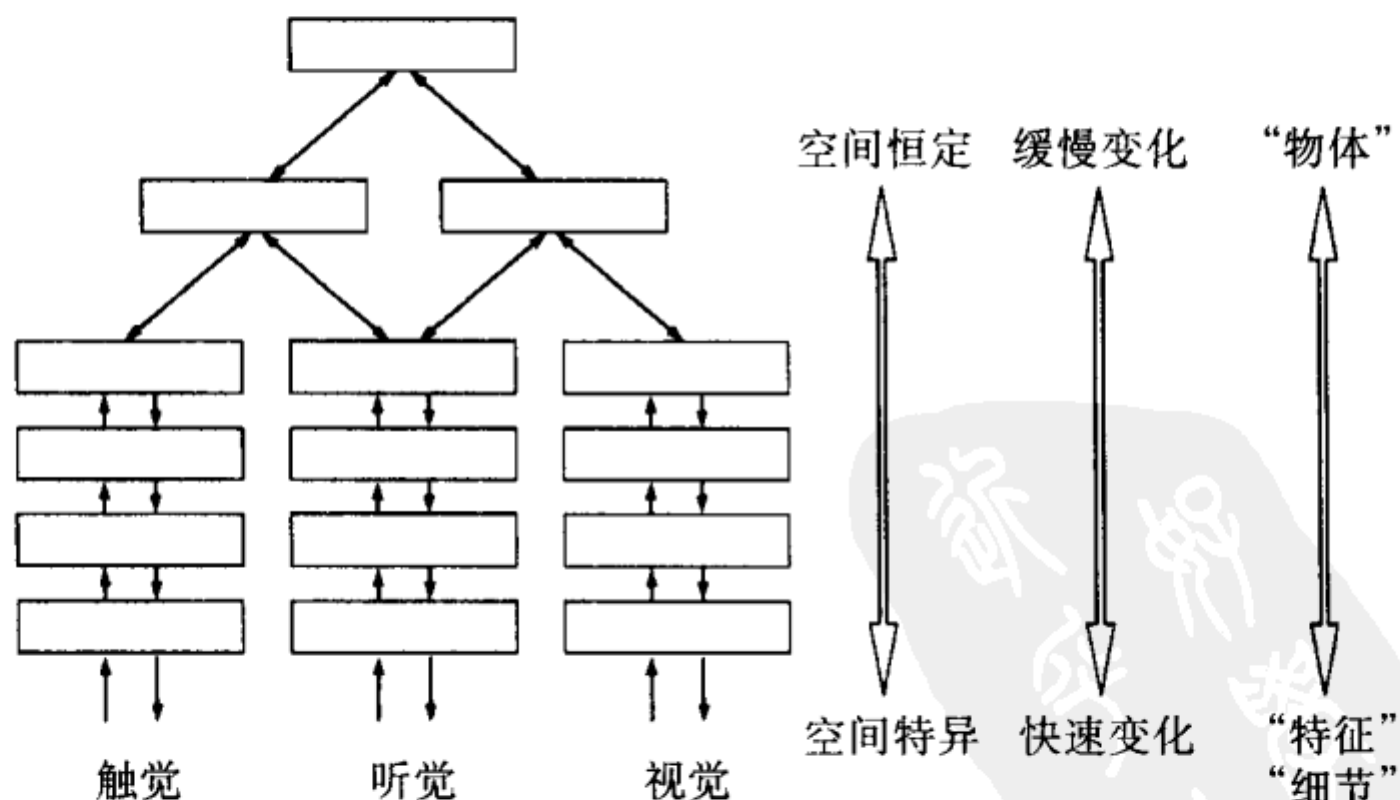


图3 形成触、听、视觉的恒定表征

这些关于视觉的从快速变化到缓慢变化、从空间特异到空间恒定的转化过程，在很多的文献中有所论述，虽然只有少量

的证据能提供证明，但许多神经科学家都认为在脑皮层的全部感觉区域里都是同样的情形，而不仅仅在视觉区域里是这样。

下面我们来看看听觉。当有人对你说话时，声压的变化很快，进入基本听觉区域——A1 区的模式也同样快速地变化。可是，如果将一根探针插入听觉的较高区域，我们会发现一些恒定不变的细胞对单词、有时甚至是短语做出反应。听觉皮层也许有一些细胞会在你听到“谢谢”时触发，而另一些细胞会在听到“早安”时触发。假定你识别了这个短语的话，这些细胞在整个句子所经历的时间内都会保持活跃。

第一个听觉区域接收到的模式可以有很大的变化，我们可以用不同的口音来说一个词，可以用不同的音高和不同的语速。但是在较高层次的脑皮层，这些低级特征将变得无关紧要，不管声学的细节多么不同，一个词就是一个词。音乐也是这样。《三只瞎老鼠》的曲子可以用钢琴弹出来，也可以用单簧管吹出来，或是孩子唱出来。在不同的条件下，A1 区接收的模式是迥然不同的。但是插入较高的听觉区域的探针会让我们发现每当听到《三只瞎老鼠》时，有些细胞会不断地被触发，而不管是什么乐器演奏的，节奏如何，或是其他许多细节有什么不同。当然，这样的实验并没有做过，因为这样的实验在人身上施行未免太不人道了。但是，如果你承认有一个通用的脑皮层算法，那么你就应该能确定这些细胞是存在的。我们知道，听觉皮层有类似视觉系统的反馈、预测以及恒定记忆。

触觉也不外乎这样。即使我们能够利用高解析度脑部成像仪对猴子进行研究，此类的实验也并未做过。当我坐在这里写这本书时，我的手上有一枝钢笔，我触到了笔帽，手指摸着笔

帽上的金属口袋夹，当我的手指移动时，从我皮肤上的触觉传感器传到脑皮层体觉区的模式在快速地变化，但我依然能感觉到那是同一枝笔。我可以一会儿用这几个手指、一会儿用另几个手指，甚至用嘴唇弄弯金属夹。这些输入彼此不同，它们到达脑皮层基本体觉区的不同位置。然而，我们的探针再一次发现距离最初输入几步之遥的皮层区域里，有些细胞稳定地对“钢笔”做出反应。只要我一直拨弄着钢笔，它们就一直保持活跃，而并不在乎我是用哪个手指或身体的哪个部位接触到它的。

在听觉和触觉方面，你并不能通过瞬间的输入来识别物体，你耳朵或皮肤的触觉传感器传来的模式，在某个时间点上，并没有足够的信息来告诉你听到或是摸到了什么。当你要感知一系列诸如旋律、话语或是“砰”的关门声等听觉模式时，或者当你要感知一枝像钢笔那样的物体时，唯一的办法是通过有一定时间跨度的输入信息流。只听一个音符是无法识别一个旋律的。同样的道理，只轻轻接触一下，你也不可能识别出那是钢笔的触感。因此，相对于像话语等事物的心理感知的神经活动，必须比单个的输入模式历时要久。这与先前的结论是一样的，也就是说，在脑皮层区域中位置越高，能看到的變化就越少。

视觉也是基于时间的输入流，它与听觉和触觉的工作方式一样。但由于我们通过一次注视就能够识别单个物体，这就把整幅图画给弄乱了。的确，这个通过一次短暂的注视就能识别空间图形的能力，许多年来将那些研究机器和动物视觉的研究人员引入了歧途——他们通常忽视了时间的关键性。虽说在实

验室的条件下人能被迫在不移动眼睛的条件下识别物体，但这并不是人们通常采用的方式。正常的视觉，比如你现在读这本书，就要求眼球不断地运动。

整合感觉

联合区又是怎样的呢？至此，我们已经看到了信息是如何在脑皮层的特定感觉区上下传递的。当向下传递的信息汇入到不断输入的信息时，就会对我们接下来要经历到什么进行预测。同样的过程也出现在不同感觉之间，也就是说，在视觉、听觉、触觉以及许多其他的感觉之间。例如，听到某种声音就会使我预感到要看到或触到什么。此刻我正在卧室写作，我家的猫“可儿”走动时，它的项圈会丁当作响。现在，我听到了这丁当声从走廊传过来，且越来越近。正是从这个听觉的输入我识别出了我的猫，于是我转过头去朝走廊看，“可儿”走了进来。我是根据“可儿”的声音而知道会看见它的。假设“可儿”并没有进来，或者是其它一个什么动物出现的话，我一定会感到吃惊。在这个例子中，首先是听觉的输入产生了对“可儿”的听觉识别。这个信息流经由听觉的体系到达联合区，正是这个联合区将视觉与听觉连接了起来。然后，这个表征回流到听觉和视觉的体系，从而产生听觉和视觉的预测。图4说明了这个例子。

此类多感觉的预测随时都在发生。我将钢笔的夹子向外扳，感觉到夹子脱离了手指时，我会预期听到夹子碰到笔帽而发出“劈啪”声。如果此时没有听到“劈啪”声的话，我会感到惊讶。我的大脑精确地预测了我何时将听到声音以及这个声

音听起来是什么样子。为了做出这个预测，信息向上流到脑皮层的体觉区并同时回流到体觉区和听觉皮层，从而能预测到将听到并感觉到一个“劈啪”声。

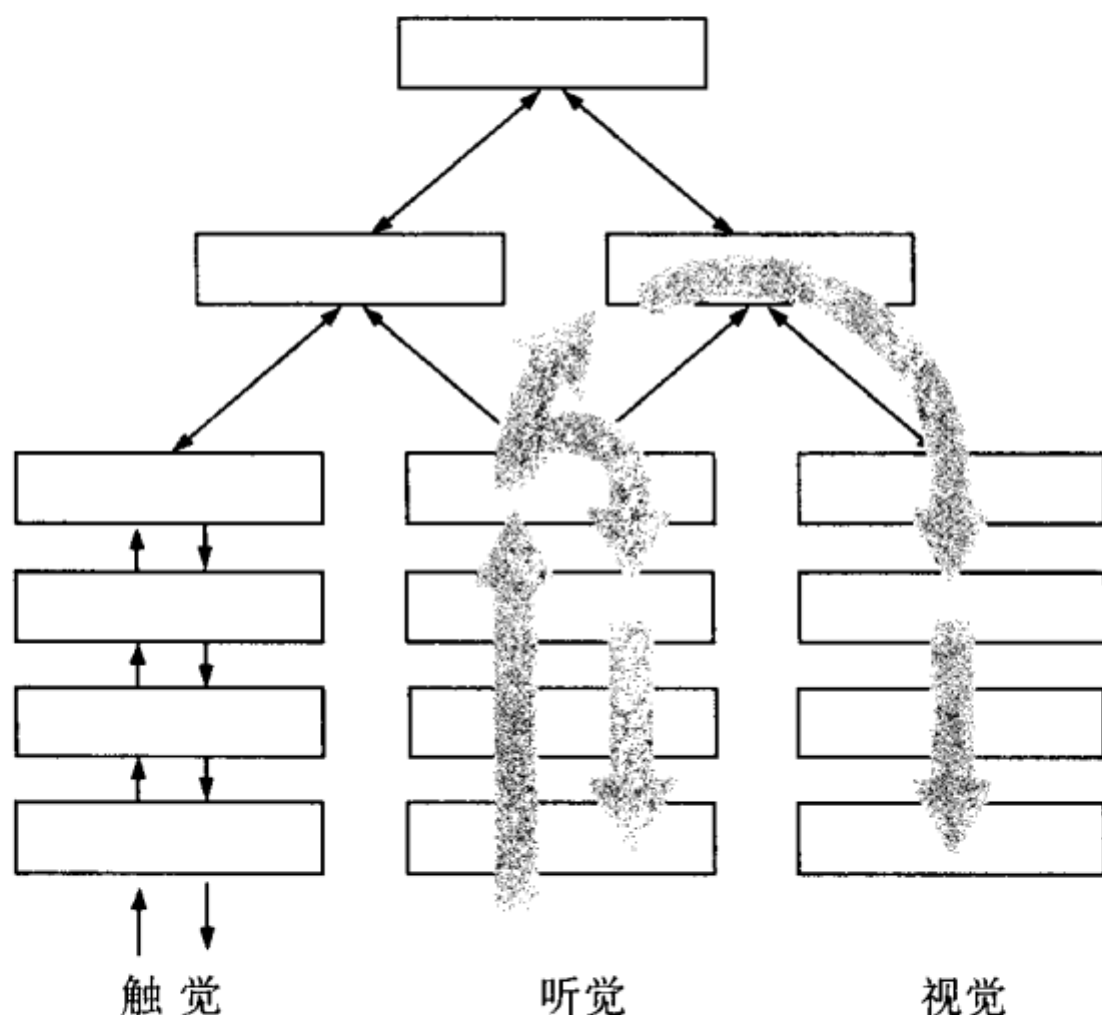


图 4 信息经感觉体系上下传递,形成预测并产生统一的感觉体验

再举一个例子：我每周有几天要骑自行车去工作。早上我走进车库，取了自行车，将它调转头来，驶上了马路。在整个过程中，我收到了许多视觉、触觉以及听觉的输入——自行车碰到了门框，链条咔哒作响，脚踏板碰了我的腿，车轮擦过地面旋转着。在将自行车从车库里拿出来的过程中，我的大脑收到了一连串视觉、听觉以及触觉的感觉，每一种感觉输入都以非常协调的方式为其他感觉做出预测。我看到的事物能准确地预测我将触摸到和听到什么。反之亦然。看到自行车撞上门框，我就会期待听到一种撞击的声音、感觉到自行车会向上跳起。当感到脚踏板碰到了腿时，我会迅速地向下看，预料到我

会在期望见到它的地方见到它。这个预测是那么的准确，以至于我会觉察出这些输入有丝毫不协调或不寻常的地方。信息在感觉体系中同时向上和向下流动，从而产生包含有各种感觉预测的统一的感受体验。

请尝试下面这个实验。放下手中的这本书，停止阅读，活动活动。活动一下身体并摆弄什么东西，比如说，走近洗手池，打开水龙头。好，当你这么做时，尽力注意每一点声音、每一个接触的感觉和每一个变化着的视觉输入。你必须集中注意力，每一个动作都与视觉、听觉以及触觉密切联系着。上提或旋开水龙头手柄时，你的大脑期待着感觉到皮肤上的压力和肌肉的阻力，你期待着看到且感觉到手柄的移动，同时也期待着看到水并听到水的声音。当水溅到水池里时，你期待着听到不同的声音，并且期待着看到和感觉到水的飞溅。

每走一步都会发出声音，不论是有意识还是无意识，你都会预期到这一点。就连手持这本书这样简单的动作也会引起许多感觉的预测。想像一下：如果你感觉到并听到书合上的声音，而你却看着它仍然打开着，你一定会惊讶万分，也会茫然不知所措。正如在第五章中改装门的那个思维实验里看到的那样，你会对世界做出各种感觉协调一致的恒常预测。当我专注于所有这些小小的感受体验时，我真为我们的知觉预测是多么的整合一致而感到惊讶。虽然这些预测看起来很简单，也不起眼，但你要记住的是，它们是无处不在的，并且它们只有在大量的模式协调一致，上、下流经脑皮层的感受体系时才可能出现。

一旦你理解了感受是怎么互相连接在一起时，你就会得出这样的结论：整个新大脑皮层及所有的感受和联合区域都行动

如一。是的，我们有一个视觉皮层，但它是一个完整的感觉系统的一部分。这个感觉系统包括视觉、听觉、触觉以及更多的感觉结合在一起，所有这些都在一个单一的多分支的体系中同时向上、向下流动。

还有一点，所有的预测都是从经验中学习的。我们期待着钢笔夹现在和将来都会发出“劈啪”声，这是因为它们过去就是这样。自行车在车库碰撞而发出“砰砰”的声音，我们都是以可预测的方式看到、感觉到并且听到的。你不是生来就具有这些知识，这些都是通过学习而得来的，这得归功于你的脑皮层有令人难以置信的大容量，从而可以记住各种模式。如果传入大脑的输入中有一贯相同的模式，脑皮层就会将它们用来预测未来的事件。

虽然图 3 和图 4 并没有描绘皮层的运动区，但是你可以把它想像成另一个分层的薄饼沓，它就像感觉的薄饼沓一样通过联合区连接到感觉系统（也许与脑皮层的体觉区有着更密切的关系，以便实现身体的动作）。这样的话，脑皮层的运动区就与感觉区域以几乎同样的方式工作着。任何一个感觉区的输入都能上流到联合区，联合区再产生一个模式向下流到皮层运动区，最终引起动作。正如视觉输入能导致模式传到脑皮层的听觉和触觉区，视觉输入也可以导致模式传到脑皮层的运动区。在前一种情况下，我们把这些向下传递的模式叫做预测。而在皮层运动区，我们把它们叫做运动指令。正如蒙卡斯尔指出的那样，皮层运动区和皮层感觉区看起来是一样的。因此，脑皮层处理向下传导的感觉预测的方法，就与它处理向下传导的运动指令的方式类似。

很快我们便会看到脑皮层中并没有单纯的感觉区和单纯的运动区。感觉模式同时传入皮层的任何一处和所有各处，然后回流到这个体系的任何一个区域，引起预测或运动。虽然皮层运动区有某些特别属性，但我们仍然可以把它认为是一个更大的分级结构存储器——预测系统的一部分。它和其他的感觉几乎一样。视觉、听觉、触觉和动作紧密地纠缠在一起。

V1 区新说

要解开脑皮层结构之谜，下一步需要从新的角度来审视脑皮层的各个区域。我们知道，脑皮层体系的较高区域形成了恒定表征，但是为什么这个重要的功能只出现在较高区域呢？根据蒙卡斯尔关于对称的概念，我开始着手研究脑皮层区域连接起来的各种不同方式。

图 1 描绘了视觉通路中的 4 个典型区域：V1 区、V2 区、V4 区和 IT 区。其中 V1 区处于整个结构的最底层，往上依次是 V2 区、V4 区和处于最顶层的 IT 区。按照传统的看法，每一个区域都被看成是单独的连续的区域，因此 V1 区的所有细胞虽说各自处理视野中的不同部分，但它们恐怕都在完成类似的任务。V2 区的所有细胞也完成同样的任务，V4 区的细胞也是一样。

首先，依照这个传统的观点，当一张脸的图像进入 V1 区时，那里的细胞建立起这张脸的一个草图，这个草图是由简单的线段和其他基本的细节所构成的。这幅草图被传到 V2 区，V2 区也对图像进行处理，对五官特征稍作精密的分析，然后将它传到 V4 区，以此类推。只有当输入信息到达顶层的 IT 区时，才产生恒定表征以及对该物体的识别。

可惜的是，这个关于 V1 区、V2 区和 V4 区等脑皮层区域的统一看法存在着某些问题。我们不禁再一次心存疑问：为什么恒定表征只出现在 IT 区？如果脑皮层的所有区域功能都一样，为什么 IT 却有所不同？

其次，脸可以出现在 V1 区的左侧，也可以出现在右侧，两种情况下你都能识别出来。但实验显示，V1 区的不相邻各块却不是直接连接起来的，V1 区的左侧并不直接知道右侧所看到的東西。请好好想想这一点。显然，V1 区各部分功能相似，因为它们都参与到脸的识别过程，但同时它们在生理上又是相互独立的。V1 区的各子区域生理上相互分离，但却在完成同样的任务。

最后，实验显示脑皮层中所有的较高区域都是从位于它们下面两个以上的感觉区域接收到整合输入的（图 3）。在人脑里，数十个区域汇聚于联合区。但根据传统的看法，V1 区、V2 区和 V4 区这样的较低层的感觉区域似乎是以不同的方式连接起来的。每一区看起来好像只有一个输入源，即只有一个由底而上的箭头，而不是明显的来自不同区域的整合输入。V2 区接受来自 V1 区的输入，仅此而已。那为什么脑皮层的某些区域接收整合输入而另外一些却不是呢？这显然与蒙卡斯尔关于脑皮层的通用算法的观点是相矛盾的。

由于上述的理由和一些其他的原因，我渐渐认为，V1 区、V2 区和 V4 区不应看成是单一的脑皮层区域。与此相反，我认为每一区都是由许多较小的子区域构成的。让我们回到先前的那个餐巾比方上——整个脑皮层的一个平化版本。比方说，我们要用笔在脑皮层“餐巾”上标记出所有的皮层功能区域的

话，显然最大的区域是 V1 区，即基本的视觉区域。其次是 V2 区。相对于多数区域来它们是非常大的。实际上我要说的是，V1 区应该被认为是由许多很小的区域组成的。在“餐巾”上原本指派给 V1 区的区域上，我们可以画出许多小的区域，而不是一整块大的区域。换言之，V1 区是由许多彼此分离的小皮层区组成的，这些小的区域通过体系中较高的区域而与周边相邻的区域间接地相连。在所有的视区当中，V1 区的子区域数量最多。V2 区也是由许多子区域组成的，只不过这些子区域数目较少，面积较大。V4 区也是如此。但顶层的 IT 区却是一个完整的单一区域，这一区域的细胞对整个视觉世界具有鸟瞰的视野。

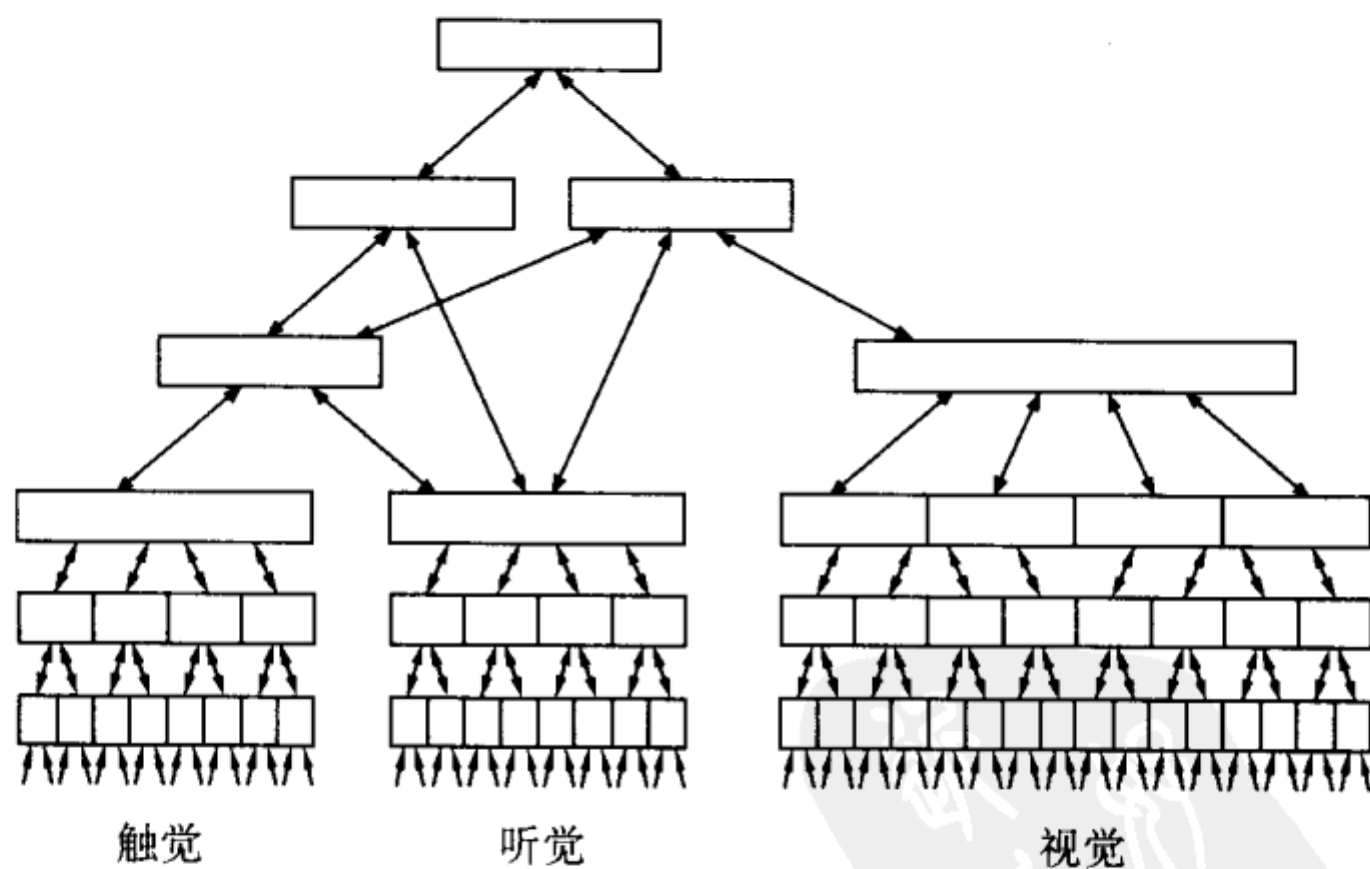


图 5 对脑皮层体系的新看法

这里有一个令人高兴的对称现象。请看一下图 5。它与图 3 描绘了同样的体系，只不过其中的感觉区是以我刚才描述的样子描绘的。请注意，现在脑皮层看上去到处都是相似的了。任意挑一个区域来看，你都会发现其下方许多子区域提供整合的

感觉输入。接收信息的区域将投影发送回到输入区，告诉它们接着将会看到什么模式。较高的联合区整合了来自诸如视觉和触觉等多个感觉区的信息，像 V2 区这样较低的区域整合来自 V1 区内部各子区域的信息。每个区域都不知道，也不可能知道这些输入都意味着什么。V2 区的一个子区域不必知道它自己正在处理来自 V1 区几个部分的视觉输入，联合区也不必知道自己在处理来自视觉和听觉区的输入。相反，任何一个脑皮层区域的工作就是要弄清它的这些输入是如何关联的，然后记忆这些输入之间的关联序列，并用这个记忆来预测这些输入将有何种表现。脑皮层就是脑皮层，各处的工作都是一样的，这就是通用的脑皮层工作原理。

这样一个新的分层描述有助于我们理解建立恒定表征的过程。下面我们以视觉为例更深入地看看它是如何工作的。在第一个层面加工过程中，视觉空间的左侧与右侧是不同的，这个不同正如听觉不同于视觉一样。V1 区的左右两侧能形成同样的表征，这是因为它们都曾接触过现实中相似的模式。正如听觉和视觉一样，它们可以被看成是彼此独立的感觉流，并在较高的层次上得到整合。

与此相似，V2 区和 V4 区内的小区域都是视觉的联合区(子区域可能有重叠，但这不会从根本上改变这些区域工作的方式)。这样来解释视觉皮层并不会与我们已知的解剖学知识相冲突，也不会改变我们对它的解剖学构造的看法。信息顺着这个分级树状记忆结构的分枝上下流动。视觉区左侧的一个模式可以引起右侧的预测，这与猫的铃铛可以引起我做出的关于它会进入我卧室的视觉预测的道理是一样的。

关于脑皮层体系的这样一种新的描述带来的最重要的好处是，现在我们可以说脑皮层的每一个区域都形成恒定表征。按照老的思路，只有等输入到达顶层，也就是能看到整体视觉世界的 IT 区时，我们才形成彻底的恒定表征，比如说“脸”的表征。而现在我们可以说恒定表征是无所不在的。在脑皮层的每一个区域都形成有恒定表征。表征的恒定性并不是在较高的皮层区域——IT 区才变戏法般地出现的。每一个区域都从位于自己下面的输入区获得并形成恒定表征。因此我们说，V4 区、V2 区和 V1 区的子区域都根据流入它们的信息产生恒定表征，它们只不过看到世界的很小一部分，它们涉及的感觉物体的符号也是非常基本的，但它们却和 IT 区一样完成相同的工作。同时，IT 区上面的联合区从多种感觉中形成模块式的恒定表征。因此，脑皮层的所有区域都产生在体系中位于自己下方的那部分的恒定表征。这是一幅漂亮的图景。

我们的问题有了变化。我们不再追问恒定表征是如何在自下而上的 4 个层级中得以产生，然而我们却要追问恒定表征在每一个脑皮层的区域中是如何产生的。如果我们确信脑皮层有一个通用的工作原理，那么这样设问是完全有意义的。如果一个区域存储模式序列，那么每一个区域都存储序列；如果一个区域产生恒定表征，那么所有的区域都产生恒定表征。沿着图 5 中所示的线条重新审视脑皮层的体系，这种解释就会更容易被理解。

世界的模型

为什么新大脑皮层具有层级结构呢？你能够思考这个世界，也可以在世界中走动，还可以对未来做出预测。所有这些都是

因为你的脑皮层构建了一个有关世界的模型。本书最重要的一个思想就是，脑皮层的层级结构存储着有关现实世界的一个层级结构模型，现实世界的嵌套结构在脑皮层的嵌套结构中得以反映。

嵌套结构或者层级结构又是什么意思呢？想想音乐吧！音符合在一起构成音程。音程合在一起构成旋律子句。旋律子句组合成旋律或歌曲。歌曲的集合就是歌曲集。再想想书面语吧。字母合在一起构成音节。音节合在一起构成单词。单词组合成子句和句子。下面把这个问题反过来看看。想想你的周围吧！你周围一定会有路、有学校和房子。房子里有房间。每个房间都有墙壁、天花板、地板、门以及一个或多个窗子。而这每一个物体都由较小的其他物体组成。窗子由玻璃、窗框、窗门和纱窗组成。窗门又由像螺钉等的更小的部分组成。

稍微花点时间抬起头看看你的周围。来自视网膜的模式进入到基本的视觉皮层，在这里，这些模式结合起来形成线条。线条又组合形成更复杂的形状。这些复杂的形状再组合形成事物，比如说鼻子。鼻子与眼睛和嘴一起构成了脸。脸又和其他的一些身体部分组合成比如说是坐在房间里你对面的一个人。

世界上的一切事物都是由同时出现的较小事物组成的。事物就是这样。当我们为一个事物命名时，我们之所以能够命名它是因为它有一组细节同时存在。脸之所以是脸，是因为两只眼睛、一个鼻子和一张嘴总是同时出现。眼睛之所以是眼睛，是因为瞳孔、虹膜、眼睑等等总是同时出现。椅子、汽车、树、公园和乡村莫不如此。最后，歌曲之所以成为歌曲是因为一系列音程总是依次连续出现。

如此，世界就像一首歌。世界上的每一个物体都是由许多较小的物体组成，而大多数物体都属于更大的物体。这也就是我所说的嵌套结构。一旦你了解了这一点，你就随处可见嵌套结构。与此极其相似，你对事物的记忆以及你的大脑表征事物的方式就存储在脑皮层的层级结构里。你对于自己家的记忆并不存在于脑皮层的某一区域里。它存储在脑皮层的整个层级结构里，而这又反映了你家的层级结构。大的关系存储在这个体系的最高层，而小的关系则存储在底部。

脑皮层的设计以及它自然的记忆方法展示了世界的层次关系。你生来并没有关于语言、房子或是音乐的知识。脑皮层有一个巧妙的学习算法，它能自然地发现存在着的任何层级结构并捕获它。当结构不复存在时，我们就陷入了一片混乱，甚至是混沌状态。

在任何一刻，你只能体验到世界的一个子集。在一个时间点上，你只能在你家里的一个房间里，看着一个方向。因为有了脑皮层的层级结构，你就能知道你是在家里，在起居室里，看着窗户，即使当时你的眼睛只注视在一个窗台上。脑皮层的较高区域保持着你家的表征，较低的区域则表征房间，更低的区域则看着窗户。同样地，这个层级结构让你能知道是在听一首歌、一个歌曲集，即使在任何一刻你只听到一个音符，而这个音符本身几乎并没有意义。它也能让你知道你是和最要好的朋友在一起，即使在那一瞬间你的眼睛只注视着她的手。脑皮层的较高区域追踪着大的图像，而较低的区域则积极地处理快速变化着的小细节。

因为在任何时刻我们只能接触、听到和看到世界很小的一

部分，传入大脑的信息自然就以一系列的模式到达大脑。脑皮层要记住这一再出现的序列。有时候，例如旋律，模式的序列是严格顺序的，比如说按音程的顺序。我们多数人都熟悉这种序列。但是我所用的序列一词则更加宽泛，比较接近数学上的“集合”的含义。一个序列是一组通常彼此伴随的模式，不一定总是有严格的顺序。重要的是一个序列中的模式在时间上是先后相继的，即使不一定按固定的顺序。

举几个例子可以使问题变得更清楚。当我看着你的脸时，我看到的输入模式的序列不是固定的，而是由我的扫视而决定的。某个时候我可能以“眼睛—眼睛—鼻子—嘴”的顺序来注视，而过了片刻我又是以“嘴—眼睛—鼻子—眼睛”的顺序注视的。组成脸的各部分就是一个序列。它们在统计上是关联的，并且总是几乎同时出现，虽然顺序有可能改变。如果你感觉到看到了一张“脸”，而此时你正注视着“鼻子”，那接下来的模式很可能就是“眼睛”或“嘴”，而不会是“钢笔”或“汽车”。

脑皮层的每个区域都看到一股这样的模式流。如果模式都是如此相互关联着，而皮层区域能够学会预测接下来将出现什么模式的话，我们就说脑皮层区域对这一序列形成了稳定的表征，或者叫记忆。产生真实世界中的物体的恒定表征的最基本要素就是要记住序列。

真实世界的物体可以是具体的，如蜥蜴、脸或门等，也可以是抽象的，如词或理论。大脑处理抽象和具体的事物的方式是一样的。这些事物不过是以可预测的方式在时间上同时出现的模式序列。一定的输入模式反复出现的事实使得脑皮层区域

知道这样的体验是由世界上的实物所引起的。

现实的基本要义就是可预测性。如果脑皮层区域发现自己能够通过一系列的身体动作(例如眼睛的扫视或者手指的抚摸)可靠地且有预见性地在这些输入模式间切换,并且当它们不断明朗化时(例如声音能组成一首歌或一个词),能够对它们做出精确的预测,大脑就认为它们之间存在因果关系。纯粹由巧合引起的大量输入模式再三出现在同一关系中的几率几乎是微乎其微。一个可预测的模式序列一定是一个实实在在存在着的更大事物的一部分。因此可靠的预测性以铁定的方式说明了世界上不同的事件实际上是相互联系着的。每一张脸都有眼睛、耳朵、嘴和鼻子。如果大脑看到了一只眼睛,然后扫视到另一只眼睛,然后又扫视到嘴,它就可以确信自己看到的是一张脸。

如果脑皮层区域能够说话,它们也许会说:“我体验到了许多不同的模式。有时候我预测不到接下来会看到什么模式,但是这些模式的集合肯定是彼此联系着的。它们总是一同出现,我能够可靠地在它们之间跳来跳去。因此每当看到这些事件中的任何一个,我都会用同一个名字来指称它们。我传递给皮层中更高区域的就是这样的组名,而不是单个的模式。”

因此,我们说大脑存储着序列的序列。脑皮层的每一区域都记住序列,给它所知道的序列命上“名字”,然后将这些名字传递给脑皮层层级中下一个更高的区域。

序列的序列

信息从基本的感觉区域流到较高层级的过程中,我们可以看到变化是越来越少的。在 V1 这样的基本视觉区里,随着新

的模式以每秒几次的频率落在视网膜上，兴奋的细胞组在不断变换。而在 IT 视觉区里，细胞的兴奋模式就相对稳定多了。这是怎么回事呢？脑皮层的每一区域都有一个已知序列的汇集，就好像歌曲汇集一样。皮层区域为任何事物都存储着诸如歌曲似的序列：海浪冲击海滩发出的声音、你母亲的面貌、从家到附近商店的路、怎样写“爆米花”这个词、怎样洗纸牌，等等。

歌曲都有自己的名字，类似地，每一个脑皮层区域都为已知的每一个序列命了名。这些“名字”就是一组细胞，它们的集体兴奋表示了序列中的这一组事物。（现在先不要关心哪组细胞是如何被挑选出来表征哪一序列的，我们在后面将会讲到。）只要这一序列仍在持续，这些细胞就会一直保持兴奋状态。传递给层级中更高区域的也就是这个“名字”。只要输入模式属于可预测的一个序列，该区域将会向下一更高层级的区域提供这个恒定的“名字”。

就好像这个区域在说：“这就是我所听到、看到、触摸到的序列的名字。你不必知道单个的音符、直线或纹理结构。一旦有新的或未曾预见的事物出现时，我就会通知你。”说得更准确些，也就是，我们能够想像处于视觉体系最高位的 IT 区向处于它之上的联合区传达到：“我看到了一张脸。眼睛的每一次扫视都注视在脸的不同部分，但我的确是见到了一张脸；我连续地看到脸的不同部分，但那就是同一张脸。一旦看到其他事物，我马上会通知你。”一个可预测的事件序列就以这种方式与一个“名字”——一个细胞兴奋的稳定模式——联系在一起了。如果我们沿着这个分层的金字塔一直向上观察的话，这

种情形会反复出现。一个区域也许能识别一个由音素(组成词的声音)组成的声音序列,并将代表音素的模式传递到下一个区域。更高层级的区域识别出音素序列并生成单词,更高层级的区域识别出单词序列并生成短语,如此往下进行。要牢记于心的是,脑皮层最底层区域中的“序列”可能相当简单,例如平移过空间的一条可见线段。

在体系中的每一个区域里将可预测的序列变成“已命名的对象”。这样,在体系中层级越高,稳定性就越强。恒定表征就这样产生了。

当模式沿着体系往下传播时,就会出现相反的结果:稳定的模式“展开”成序列。我们现在假设,你在7年级时已背熟了葛底斯堡演讲,而你现在想将它再背出来。在处于脑皮层高层的语言区域里,存储着一个表征着林肯这一著名演说的模式。首先,这个模式展开成一个记忆中短语序列。在下面的一个区域中,每一短语又被展开成单词序列。而就在此刻,展开的模式分成两支,分别沿着脑皮层的听觉区和运动区向下传递。在运动区的那条路径中,每个词又被展开成音素序列。到了最后,也就是在最底层,每个音素被展开成肌肉指令序列以发出声音。这个体系中越是往下,模式变化得越快。处于运动区域最高层的那个稳定的模式最终产生了一个复杂而绵长的语音序列。

当信息在系统内部下行时,恒定性也在起作用。当你想将葛底斯堡演讲用打字机打出来而不是念出来时,在系统中的最高层级中是以相同的模式开始的。在下面的区域中该模式被展开成短语。在更下一层的区域中短语又被展开成单词。到目前

为止，念出葛底斯堡演讲和打出葛底斯堡演讲之间并没有任何差异。但在随后的下一层级中，脑皮层运动区却走了另一条路径。这次，单词被展开成字母，字母又被展开成发给手指肌肉的打字指令。“Fourscore and seven years ago our fathers brought forth...” (87年前，我们先辈在这个大陆上创立了一个新国家……) 这些单词的记忆被处理成恒定的表征；这与你是念出来，还是用打字机打出来，还是用手写出来没有关系。请注意，你不必将演讲存上两份，一次是为念的，一次是为写的。这个演讲的单份记忆就能产生多种行为。在任何一个区域里，恒定模式都能分为两支，分别沿着不同的路径向下传播。

系统中最底层的简单对象的表征能够被各种更高级的序列反复利用，这也可以说是比较有效的方式。例如，葛底斯堡演讲和小马丁·路德·金(Martin Luther King Jr.)的“我有一个梦想(I Have a Dream)”的演讲中的单个单词你不必记上两套，即使这两个演讲中包含了一些相同的用词。一个由嵌套序列构成的体系使得我们能共享、重复利用低级层级所存储的事物——单词、音素和字母，等等。这是一个非常有效的存储世界及其结构的方法，它与电脑的工作方式是决然不同的。

序列的这种展开方式也出现在感觉区，而不仅仅是出现在运动区，展开的过程也能让你从不同的方面去感受并理解事物。如果你走近冰箱去取冰激凌，那么你的视觉皮层会在多个层级上变得兴奋起来。在比较高的层级上，你感觉到一个恒常稳定的“冰箱”。在较低的区域里，这个视觉上的预期分解成一系列更多的局部视觉输入。看到冰箱这一行为是由注视冰箱的门把手、造冰机、门上的磁吸、孩子画的画等组成的。就在

你从冰箱的一个局部特征扫视到另一个特征的几毫秒的瞬间，对每一次扫视结果的预测也同时在视觉体系中向下传播。只要这些预测随着一次次的扫视得到确认，高层的视觉区域就确信你看到的就是个冰箱。请注意，在这一例子中，与葛底斯堡演讲中词的固定顺序有所不同的是，看着冰箱时所感知到的序列不是固定的——视觉输入流与提取的记忆模式流取决于你动作的顺序。因此在此例中，展开的模式不是严格次序的序列，但最终结果却是一样的：缓慢改变的高级模式展开成快速变化的低级模式。

当信息在脑皮层体系中上、下流动时你的记忆序列以及用名字表征它们的方式可能使你联想起军令的体系。司令官发令道：“调兵佛罗里达过冬。”随着命令在系统中向下传达，这一简单的高级命令被展开成一系列许多详细的指令。司令官的部下明白这一命令要分解成许多的步骤，比如说准备出发、将部队往佛罗里达转移以及准备到达等。这些步骤又被进一步分解成更小、更具体的步骤，由下一级来完成。到了最底层，上千士兵完成成千上万的动作，部队才得以行进，然后逐级形成命令实施的情况报告。随着这些报告向上呈报，它们被反复地汇总，直到到达顶层，最后司令官收到一份日志：“调兵佛罗里达完毕。”他不需要了解所有的细节。

但这一规律有一种例外情况。如果发生了事故，命令链上的下属无法处理，那么问题就会被逐级提交直到有人知道怎么处理为止。知道如何处理该问题的军官并不把它看作是例外。对部下来说，没有预测到的问题却是他规划中已预计到的下一个任务：军官向下属重新发出新的指令。新大脑皮层的工作方

式就与此类似。我们很快就会看到，当没有预期到的事件(或模式)出现时，关于它们的信息就会在脑皮层系统中向上传递，直到有一个区域能够处理这些信息。如果脑皮层中较低的区域无法预测它们看到的是什么模式，它们就会认为这是一个错误并且将这一错误向上传递。这一步骤反复进行直到某一区域确能预测出这一模式。

无一例外地，脑皮层的每一区域都在努力存储并且提取序列。但这样来描绘大脑的工作原理仍然过于简单。我们要在这一模型中增加稍许的复杂性。

这些到达脑皮层某一区域的“自下而上”的输入是由成千上万以至上百万的轴突所承载的输入模式形成的。这些轴突来自不同的区域，且包含着各种模式。即使是在 1 000 个轴突上可能的模式数量也比宇宙中的分子数还要多。在人的一生中，一个脑皮层区域只能看到这些模式中的极小一部分。

那么，问题就出来了：当一个区域存储序列时，这些序列又是关于什么的序列呢？问题的答案就是：这个区域先将输入分类，将它们归到有限的几种可能性中，然后再寻找能与之对应的序列。假设你就是一个脑皮层区域，任务是分拣彩色纸片。要将纸片分拣到 10 个桶中，每个桶上都贴有一张色标样片，其中一个桶装绿色纸片，一个装黄色，另一个装红色，等等。然后，把彩色纸片一张一张递给你，你必须将它们按颜色分类。递给你的每张纸片都稍有不同，因为世界上的颜色是无穷尽的，因此你不会拿到颜色完全相同的两张纸片。有时你会比较容易

确定一张纸片该放进哪个桶中，而有时却会有些困难。一张纸片的颜色有可能介乎红色和橙色之间，但你必须做出选择，要么放进红色的桶，要么放进橙色的桶，哪怕判断可能是随机的。（举这样一个例子是要说明大脑必须将模式进行分类。脑皮层的各区域都要完成类似的工作，当然不是将模式放入桶里。）

然后，你的新任务是寻找序列。你发现“红—红—绿—紫—橙—绿”经常共现。请注意，如果纸片没有事先被分成不同的类别，那么要识别序列便是不可能的。在没有将每张纸片进行分类的前提下，你无法确定两个序列是否完全一样。

现在你就可以开始了。你将看到所有的输入模式——从低层的皮层区域传上来的彩色纸片——将它们归类，发现序列。将信息分类和序列化是产生恒定表征所必需的两个步骤，也是皮层中每个区域都要进行的两个步骤。

当一个输入有歧义时，比如一张纸片的颜色介乎红色和橙色之间，形成序列的这个过程就能派上用场了。即使你不能确定它到底是红色多一点还是橙色多一点，你也不得不为该纸片选一个桶来装。如果你知道与输入对应的最有可能的序列是什么的话，那就可以利用这个知识来决定如何将歧义的输入进行分类。假如你因为刚刚见过两张红色的、一张绿色的和一张紫色的纸片，从而认为当前的序列是“红—红—绿—紫—橙—绿”，那么你就会预期下一张会是橙色的。但是，当下一张纸片到你手上时，你却发现它不是橙色的。它是介乎红色和橙色之间的一种颜色，看起来好像更接近红色。但你更熟悉“红—红—绿—紫—橙—绿”这样的序列，并也正期待着这样的序

列。所以你会把这张纸片放进装橙色纸片的桶。这时你就是利用了已知序列的前后关系来消除歧义的。

在我们的日常体验中时刻都能遇到这样的现象。当人们说话时，脱离了具体上下文，他们所说的单个词语就可能让人听不懂。然而，当你听到一个句子中的某个模糊不清的词时，不会因为这个词而使整个理解过程受阻，你照样能理解这个句子。同样地，离开了上下文，手写的字也经常有认不出的时候，但是，如果将它放到一个完全手写的句子中，你就能认出来了。在多半情况下，你是意识不到自己在利用大脑中的序列记忆来填充模棱两可或不完整的信息的。你听到了你所期待听到的，看到了所期待看到的——至少当你所听到的和所看到的符合过去经验时是这样。

请注意，序列记忆不但能让你消除当前输入中的歧义，还能帮你预测接下来的输入会是什么。当你的大脑皮层在分拣颜色纸片时，你会对递给你纸片的人说：“嘿，万一你要是决定不了接下来要递给我什么，根据我的记忆，应该是橙色的了。”识别出一个模式序列后，大脑皮层区域就会预测接下来的输入模式会是什么，并告诉下面的区域该期待什么。

皮层区域不但识记熟悉的序列，还学习如何对其已有分类进行修改。比如说，你一开始有了一些贴有“绿”“黄”“红”

“紫”和“橙”字样的桶，你准备要识别“红—红—绿—紫—橙—绿”序列或由这些颜色组成的其他组合，但如果某种颜色呈现出很大的不同，那又怎么办呢？如果每次你都能识别出“红—红—绿—紫—橙—绿”序列，而紫色总有些不太像紫色，那又会怎么样呢？这种不太像紫色的颜色更像是靛蓝。因此，你把“紫

色”桶的标签改为“靛蓝”，这时，这些桶才更符合你所看到的，你也会减少产生歧义的可能性。可见，大脑皮层是灵活的。

在大脑皮层的各区域中，“自下而上”的分类和“自上而下”的序列化总是相互作用，它们在你的一生中总在变化。这就是学习的本质。实际上，皮层的所有区域都有“可塑性”——它们都可以因经验而改变。形成新的分类和新的序列就是你存储世界的方式。

最后，我们来看一下这些分类和预测是如何与更高层级的区域进行交互的。皮层工作的另一方面就是將你所看到的序列的名字传递给上一层区域，也就是说你将写有“红”“红”“绿”“紫”“橙”“绿”等字样的纸片传给上一层级，它们对上方的区域来说本身并没有多大意义，因为“名字”不过是一个模式，但它们将与其他经过分类的输入结合成更高层级的序列。与你刚才所做的一样，这个上方的区域也在跟踪所看到的序列。在某一刻，它可能会对你说：“嘿，万一你无法确定下面要传给我什么，根据我的记忆，我预测应该是‘黄—黄—红—绿—黄’序列了。”从基本上讲，这是指导你应该在自己的输入流中寻找什么的指令。你将尽可能把所看到的内容解释成那个序列中的一部分。

既然许多人都听过人工智能和机器视觉研究领域中的一个术语——模式分类，那就让我们来看看模式分类这一过程与大脑皮层的处理有何不同。为了让机器能识别物体，研究人员一般都会建立一个模板，比如，杯子的图像或者杯子的原型，然后让机器学着将输入与杯子的原型进行匹配。如果两者在很大

程度上匹配，计算机就会说它看到了一个杯子。但我们的大脑中并没有类似的模板，而且每个皮层区域作为输入所接受的模式也不是图像。你所记住的并不是视网膜所看到的事物的快照，也不是耳蜗或皮肤所接受的模式的快照。大脑皮层的层级结构保证了对事物的记忆分布于皮层的层级结构中，而并不集中存于一点。并且，因为在皮层体系中每一个区域都形成恒定的记忆，所以，典型的皮层区域所学得的就是恒定表征序列，这些序列本身也是恒定记忆的序列。人脑中是找不到杯子或其他物体的图像的。

与照相机的记忆所不同的是，你的大脑记住的是世界的本质，而不是它看上去的那个样子。当你对世界进行思考时，你是在提取那些对应于世界上物体的存在和表现方式的模式，而不是它们在某一刻通过你的任何感官所表现出来的样子。你用以体验世界万物的序列反映的就是世界的恒定结构。你体验世界各部分的顺序是由世界结构所决定的。比如，你是通过直接走下登机走廊，而不是从售票台登上飞机的。你赖以体验世界的序列就是世界的真正结构，也就是大脑皮层所要记忆的内容。

但不要忘了，大脑皮层里任一区域中的恒定表征，都能通过将模式沿皮层体系向下传播而被转化为你感官将要感知的样子的细节性预测。类似地，皮层运动区的恒定表征也能通过将模式沿运动体系向下传播而被转化为细节性的、情境特异的 (situation-specific) 运动指令。

大脑皮层区域会是什么样子呢？

让我们再来看看图 5 中的小方格所示的大脑皮层区域，现

在，图 6 为我们提供了有关这些皮层区域的更多细节。我们将从描述皮层区域的结构开始，让你了解皮层区域中的细胞是如何学习并提取模式序列的，以及什么是形成恒定表征和进行预测的最必要的元素。大脑皮层的各区域大小差异显著，其中最大的要数基本感觉区。比如说 V1 区，从它在大脑后部所占的空间来说，大约有一个护照那样大小，但是，正如我前面所讲过的，它是由许多只有像 BB 弹大小的微小区域所组成的。现在我们先假定典型的皮层区的大小为一枚硬币那么大。回想一下我在第三章说过的那 6 张名片，其中每一张都代表着皮层组织中的一个层级。我们为什么会说大脑皮层有层级呢？如果把

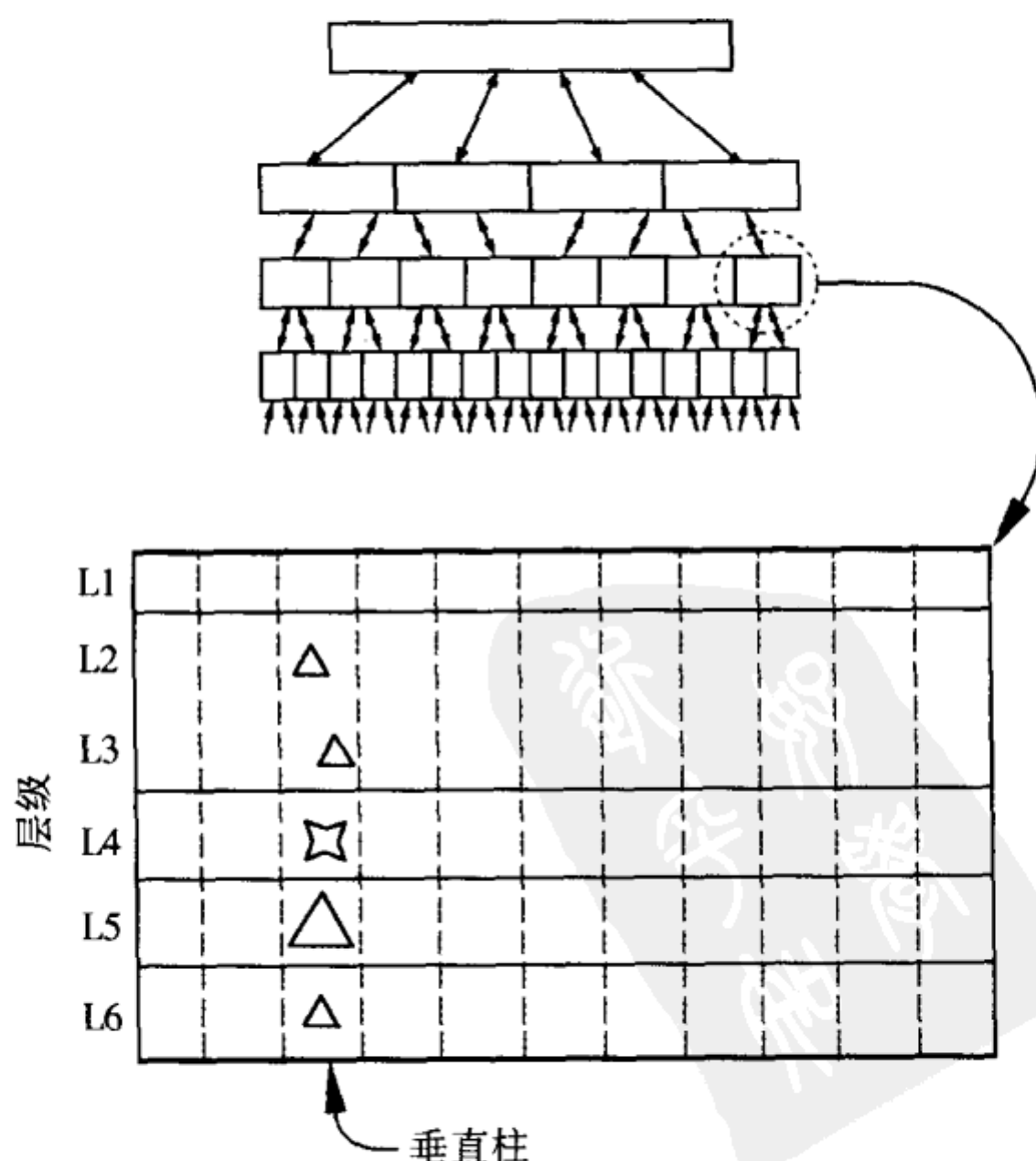


图 6 层级和垂直柱

硬币大小的皮层区放到显微镜下去观察，你就会发现从上到下细胞排列的密度和形状都不一样。我们正是根据这些差异来进行分层的。最顶层，是6层当中最清楚的一层，我们把它叫做第1层。这一层中的细胞很少，主要是由与皮层表面平行延展的轴突所组成。第2层和第3层看起来很相似，它们是由许多紧密排列的、被叫做“锥体细胞”的神经元组成，而这些神经元之所以得此名，是因为它们的细胞体很像一个小小的金字塔。在第4层中有一种星形的细胞。第5层中除了有正常大小的锥体细胞外，还有一类巨大的金字塔形状的细胞。在最底层，也就是第6层，也有几类独特的神经元。我们正是根据这些差异以及其他一些方面的不同进行分层的。

一般情况下，我们注意的是水平层，但科学家经常会谈及细胞垂直柱，这些细胞垂直柱是垂直贯穿各层的。可以把细胞垂直柱看作是由协同工作的细胞所组成的垂直单位。（“垂直柱”这一叫法在神经科学圈内有很大的争议。关于它们的大小、功能和重要性都存在不同见解。但是，就本书的目的而言，我们可以把它们看成是一种柱状的结构，这是大家都赞同的看法。）垂直柱内的各层是通过上下延展的轴突连接起来的。这样，形成了纵向的突触。这些垂直柱并不是界限分明、整齐的小柱子——大脑皮层不会这么简单——但有若干证据可以推断它们是存在的。

上述理由之一就是每个垂直柱中垂直排列的细胞对同一刺激通常会同时产生兴奋。如果我们近距离去观察V1区中的垂直柱，就会发现有些垂直柱会对朝一个方向倾斜的线段(/)产生反应，而另一些垂直柱则对朝另一个方向倾斜的线段(\)产

生反应。每一垂直柱中的细胞彼此间都有很强的联系，这就是为什么整列的细胞会对同一刺激产生反应的原因。具体地说，第4层的一个被激活的细胞会使其上的第3层和第2层中的细胞兴奋起来，然后引起底下第5层和第6层中的细胞产生兴奋。兴奋状态就在细胞垂直柱中上下传播。

我们把细胞看成呈垂直柱状的另一个理由是基于大脑皮层的形成方式。在胚胎期，单个的前身细胞从内脑室移动到皮层最终成形的地方。这些细胞中的每一个都分裂形成大约100个神经元，我们把它叫做微型柱(microcolumn)，与我刚才所描述的垂直柱连接起来。“垂直柱”一词经常被用来粗略地描述不同的现象，它可以用来指一般的垂直连通性，也可以指来自同一前身细胞的某一组细胞。根据第二种提法，我们可以说人类的大脑皮层大约有好几亿个微型柱。

为了让你对这种柱状结构有个形象的了解，请设想一下单个微型柱的粗细和我们的头发一样。我们将上千根头发扎成一束，再切成小段，切的长度大概就是不带点的小写字母i那么长。然后将这小段头发排整齐，用胶将它们粘起来，这样看上去就像一支很浓密的刷子。然后，将长头发横向粘合成薄薄的一层，表示第1层的轴突，再将其粘到刚才那块短头发上面。这个刷子状的头发块就是硬币大小的皮层区的一个简单模型。信息是顺着头发的方向流动，在第1层中水平流动，在第2层到第5层中纵向流动。

关于这些垂直柱你还需要了解另外一个细节，然后我们才能开始谈它们的作用。如果我们仔细观察的话就会发现，每一个垂直柱中，细胞上的突触有超过90%是与外部相连的，这

些突触中有一些是来自邻近的垂直柱，其他的则来自半个大脑距离以外的地方。既然大脑皮层连接中有如此大面积的横向连接，那么为什么我们要强调垂直柱的重要性呢？

答案就在于记忆 - 预测模型。早在 1979 年，当蒙卡斯尔提出单一的皮层算法时，他就指出了皮层垂直柱是皮层中的基本计算单位。然而，那时他尚不知道垂直柱都起了些什么作用。我认为垂直柱是预测的基本单位。一个垂直柱要兴奋并能进行预测，就必须知道其他地方都发生了什么，因此，就有了从各处而来的突触。

我们很快就会涉及许多细节问题，这里仅仅需要说明一下我们的大脑中为什么需要这样的连接。要预言一首歌中下一个音符是什么，你必须知道这首歌的名字，现在演奏到了歌曲中的什么地方，上一个音符结束有多久了，以及上一个音符是什么。而将一个垂直柱中的细胞连接到大脑中其他地方的大量突触，能为每一个垂直柱提供它在许多不同情境下预测下一步兴奋状态必要的上下文信息。

接下来我们要考虑的是，这些硬币大小的皮层区域（以及它们内部的垂直柱）是如何沿着皮层体系上、下发送、接收信息的。我们先来看看向上的信息流，见图 7 所示。向上的信息流沿着一条直接的路径进行传递。设想一下，我们正在观察一个有着成千上万个垂直柱的皮层区域，将这个区域进行放大，由下面区域汇流而来的输入总是先到达第 4 层——主要输入层，顺带形成了第 6 层中的连接（我们后面会谈到为什么这样做很重

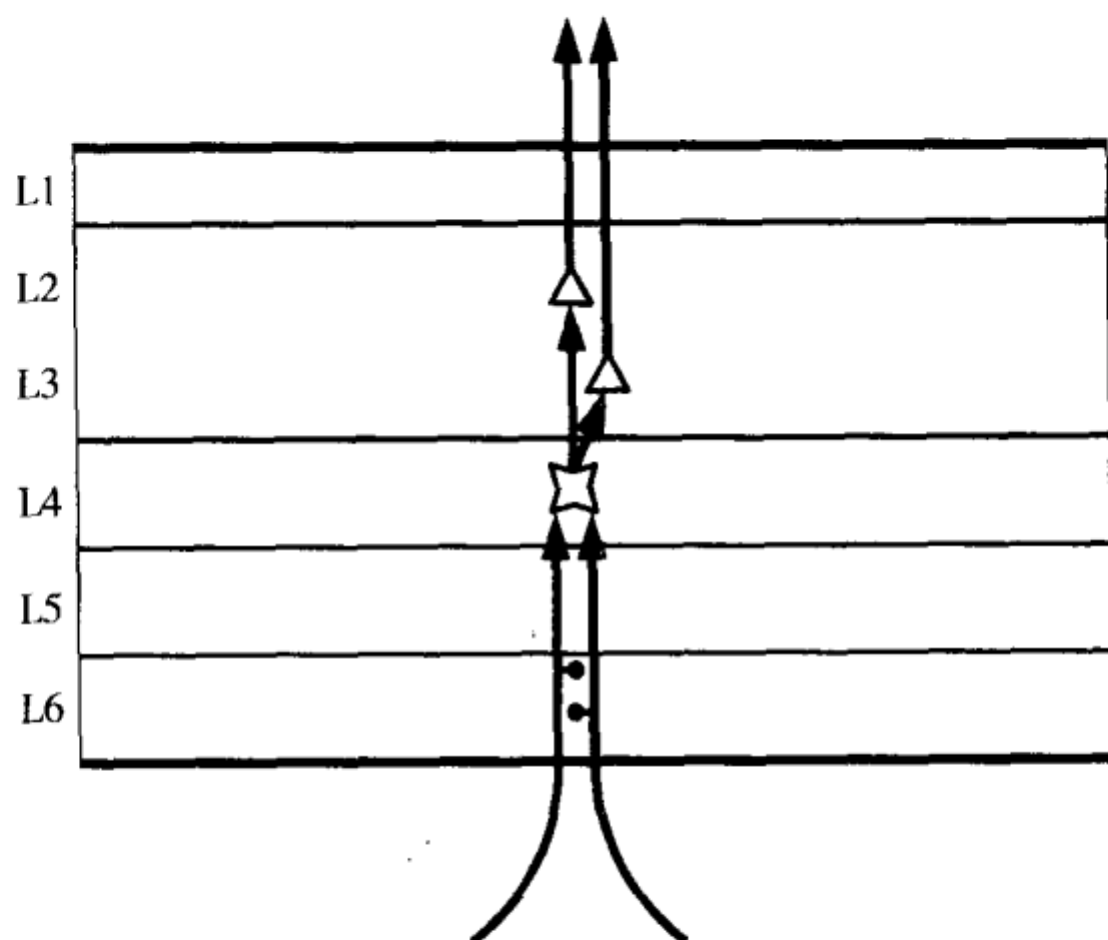


图7 在某个皮层区域中向上的信息流

要)，然后第4层的细胞在各自的垂直柱内将信息向上投射到第2层和第3层中的细胞。当一个垂直柱向上投射信息时，第2层和第3层中的许多细胞会向更高层级区域的输入层伸展轴突，这样，信息就沿着体系中的一个区域向上流动。

相比之下，皮层体系中向下信息流的路径就没有那么直接了，这可以从图8中看出来。第6层中的细胞是皮层垂直柱中向下投射的输出细胞，它们向位于自己所属区域下方的区域的第1层投射信息。在第1层中，轴突在下面的皮层区域中延展很长的距离，因此在体系中，从一个垂直柱出发的向下信息流有激活下面区域中许多垂直柱的潜能。第1层中的细胞很少，但第2层、第3层和第5层的细胞都在第1层中有树突，所以，这些细胞能被流经第1层的反馈所激活。来自第2层和第3层细胞的轴突在离开皮层的过程中会在第5层中形成突触，

从而人们认为它们能激活第 5 层和第 6 层的细胞。因此我们说，在皮层体系中，向下的信息流所流经的路径不是那么直接的。向下的信息流能通过第 1 层的传播，向许多不同的方向分流。反馈信息起源于较高层区域的第 6 层细胞，然后经由较低区域的第 1 层向四处传播。较低区域的第 2 层、第 3 层和第 5 层中的一些细胞被激活，这些细胞中有一些又激活第 6 层中的细胞，激活后的第 6 层细胞又投射到体系中较低区域的第 1 层，如此反复。（借助图 8 可以使整个过程更容易理解。）

这里，我们先来看看为什么信息的传播会遍及第 1 层。将恒定表征转化为具体预测，需要有随时决定信息在传播过程中走哪条路径的能力。第 1 层提供了将恒定表征转化为更细节化

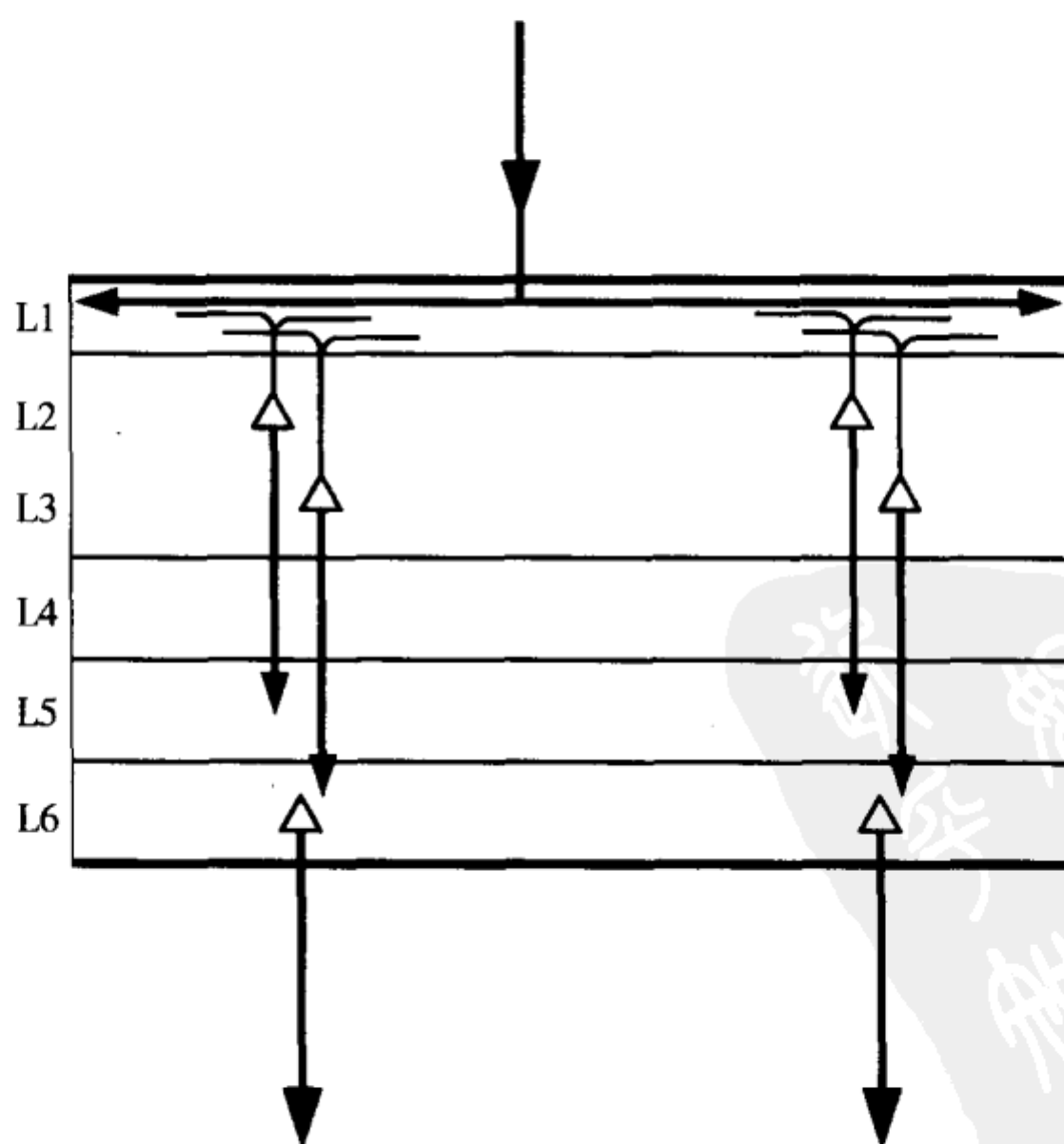


图 8 在某个皮层区域中向下的信息流

的特定表征的途径。如果你既能用口语形式又能以书面形式来回忆葛底斯堡演讲，那么口头和书面两种形式就是在共用一个表征，只是在实际调用时沿两条路径传播。同样地，当我听一首曲子中的下一个音符时，我的大脑首先取一个通用的音程，比如一个五度音程，然后将它转换成正确的特定音符，比如 C 或 G。遍及第 1 层水平流动的兴奋就提供了相应的机制。高级的恒定预测要沿皮层向下传播而成为具体预测，我们就是需要一种能允许模式流在各层级分流的机制，而第 1 层刚好能胜任这一任务。即使我们不知道这一机制到底存在与否，但我们还是可以预测出它的必要性。

最后是解剖学的发现：当轴突离开第 6 层到达其他地方时，它们都被包裹在一种叫做髓磷脂的白色脂肪物质里。这种所谓的白质就像房间里电线的绝缘层，它有助于信号间不致相互干扰，并能使信号传递得更快。信号是以大约每小时 320 千米的速度传播的。当轴突离开白质时，就进入了第 6 层中新的皮层垂直柱。

最后，皮层区域间还有另一种通信的间接方式。

描述细节之前，我们先回忆一下第二章中曾讨论过的自 - 联想记忆。你应该还记得自 - 联想记忆能用于存储模式序列吧？当一组人造神经元的输出反馈成为所有神经元的输入，并在反馈上还增加一点延迟时，模式就会一个接一个地形成序列。我认为，大脑皮层是运用了相同的机制来存储序列的，虽然可能还有一些其他的招数，但它不是由人造神经元形成自 - 联想记

忆，而是由皮层垂直柱形成自 - 联想记忆的。所有序列的输出又反馈回第 1 层，这样一来，第 1 层就知道皮层区域中哪些垂直柱是兴奋的。

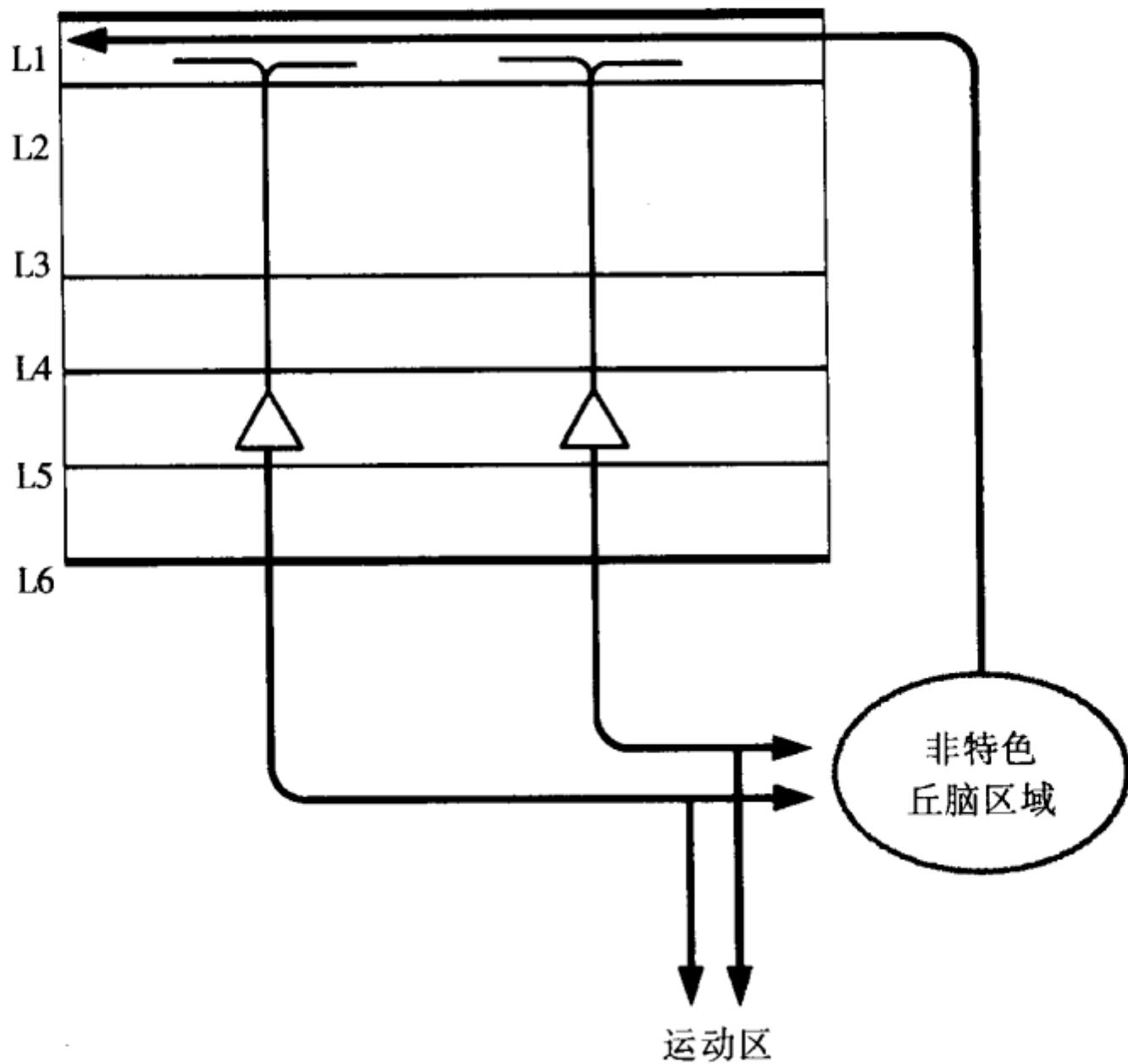


图 9 运动行为中巨大的第 5 层细胞的作用

现在让我们再来看一些细节的问题，见图 9 所示。多年来，人们已知皮层运动区(M1 区)中巨大的第 5 层细胞与肌肉和脊髓中的运动区有直接的联系，这些细胞直接驱动肌肉，使你能够运动。只要你在说话、打字或进行一些高级的行为，这些细胞就会高度协调地切换兴奋状态，使肌肉产生收缩。

近来，研究人员发现巨大的第 5 层细胞不仅在运动区，可能还在皮层的其他区域的行为中起着作用。比如，视觉皮层中

大型的第 5 层细胞会将信息投射到大脑中负责移动眼球的区域，因此，皮层的视觉区，如 V2 区和 V4 区，不但加工视觉输入，还会协助控制眼球的运动，因而你就能看到想看的东西了。巨大的第 5 层细胞在新皮层的每一个区域都有分布，并在各种运动中扮演着广泛的角色。

这些大型的第 5 层细胞不但有运动行为方面的作用，它们的轴突还一分为二，其中一支延伸到大脑中叫做丘脑的那部分，即图 9 中的圆圈部分。人类丘脑的形状及大小就像两只小鸟蛋，它位于大脑的正中间，在古脑的上方，由白质和皮层包围。皮层的每一部分都有轴突延伸到丘脑，丘脑也有轴突延伸回那些部分。对于这些连接的许多细节我们都已知晓，但丘脑本身却是一个复杂的结构，对它的作用我们还丝毫都不清楚。但丘脑对于维持正常人的生活却是相当重要的——如果丘脑受损，就会导致永久的植物人状态。

丘脑与皮层连接的路径有几条，但现在我们只对其中的一条感兴趣。这条路径从巨大的第 5 层细胞开始，这些细胞将信息投射到丘脑中的那类被认为是非特定的细胞上，这些非特定的细胞又将信息通过轴突投射回许多皮层区域的第 1 层。比如，V2 区和 V4 区内所有的第 5 层细胞都有轴突延伸到丘脑，丘脑又将信息回馈给 V2 区和 V4 区内所有的第 1 层细胞。皮层的其他部分也是如此：皮层中许多区域的第 5 层细胞都与丘脑相连，丘脑又将信息回馈到那些区域或与其相关的区域。我认为，这一回路酷似自 - 联想记忆模型中能学习序列的延迟反馈。

我们知道，第 1 层细胞有两种输入：皮层中较高区域将兴

奋传播到下面区域的第1层，某一区域内兴奋的垂直柱也通过丘脑将兴奋传播给同一区域的第1层。我们可以把这些到达第1层的输入看成是歌曲的名字(由上面区域传来的输入)和歌曲当中的位置(同一区域中兴奋垂直柱传来的延迟反馈)。因此，第1层中承载着大量信息，我们需要用这些信息来预测一个垂直柱何时应该兴奋——序列的名字以及我们处于序列中的位置。利用第1层的这两种信息，任何大脑皮层区域都可以识记与调用许多不同的模式序列。

皮层区域的详细工作原理

有了这3种回路——沿皮层体系而上的汇聚模式、沿皮层体系而下的扩散模式、经由丘脑的延迟反馈，我们就可以开始讨论皮层区域是如何完成它所具有的功能了。这里，我们想了解一下以下几个问题：

1. 皮层区域是如何将输入进行分类的(如前面的彩色纸片桶)?
2. 皮层区域是如何识记模式序列的(如曲调中的音程和面孔识别中的“眼—鼻—眼”序列)?
3. 皮层区域是如何形成稳定模式或序列的“名字”的?
4. 皮层区域是如何做出具体预测的(什么时候接火车合适，或在一个曲调中预测特定的音符)?

我们先假设皮层区域中的垂直柱就像我们用以将彩色纸片分类的桶，每一个垂直柱就相当于一个桶上的标签。每一个垂直柱中的第4层细胞接受来自下面若干区域的输入纤维，如果这些输入的组合是正确的，它就会兴奋起来。这个第4层的细

胞兴奋起来，就表明了输入与其标签相符。正如在纸片分类的比喻中输入有时可能会有歧义一样，针对同一输入可能会有多个垂直柱与之相符。这时，我们要求皮层区域做出唯一的解释：纸片要么是红色的，要么是橙色的，不能两者皆是。一个有着强输入的垂直柱应该能够抑制其他垂直柱的兴奋。

大脑中有起抑制作用的细胞，它们能完成这项工作。这些细胞能有效地抑制皮层某一区域附近的其他神经元，而只允许一个神经元兴奋，它也只能影响到垂直柱附近的范围。因此，即使有大量的抑制，一个区域中的许多垂直柱仍能同时兴奋起来。（在人脑里，没有任何事物是由单个神经元或单个垂直柱表征的。）为了容易理解起见，我们权且认为一个区域只有一个垂直柱胜出。但你要记住，实际上可以有許多垂直柱同时兴奋。皮层中一个区域将输入分类的实际过程以及它是如何学会这么做的，这个问题非常复杂，我们还无法完全了解。我不会试图让你绕开这些问题，相反，我认为皮层区域是将输入归成一组垂直柱的兴奋状态。

接下来，我们来集中谈谈序列的形成以及序列的命名吧。

皮层区域是如何存储这些分类模式序列的呢？这个问题实际上我已经回答过了，但现在我还想多谈谈有关的一些细节问题。假设你是一个细胞垂直柱，来自下面区域的输入使你的第4层细胞兴奋起来，对此你很满意，接着第4层细胞又激活了第2层、第3层细胞，然后是第5层、第6层细胞也被激活。整个垂直柱由下方驱动而变得兴奋起来。第2层、第3层和第5层的细胞都有成千上万的突触伸展到第1层，如果当第2层、第3层和第5层的细胞兴奋起来时，这些突触中有一些也

跟着兴奋起来的话，这些突触就得到了加强。如果经常这样，这些第 1 层的突触就会变得很强，足以让第 2 层、第 3 层和第 5 层的细胞在第 4 层细胞没有产生兴奋的情况下也被激活——这意味着这个垂直柱的某些部分在没有从皮层中较低区域接受输入的情况下也能兴奋起来。这样，第 2 层、第 3 层、第 5 层的细胞便学会了根据第 1 层的模式来预期它们应该在何时产生兴奋。也就是说，在这种学习之前，这个垂直柱只有在第 4 层细胞的驱动下才会兴奋起来，而在学习以后，垂直柱能通过记忆产生部分的兴奋。当垂直柱通过第 1 层的突触而兴奋起来时，它在期待来自下面的驱动。这就是预测。如果垂直柱会说话，它就会说：“过去当我被激活时，第 1 层的某些特定的突触也是被激活的。因此，当我再次看到它们兴奋时，我就会因有所预期而兴奋。”记得我们说过，第 1 层中有一半的输入是来自皮层的邻近区域或垂直柱中的第 5 层细胞，这些信息代表着片刻前曾发生过的事件，代表着在你之前已经兴奋的垂直柱。它就是曲子中的前面那个音程，或者是最后看到的東西，或是最后感觉到的东西，或者是所听话语的最后一个音素。如果这些模式在时间上发生的顺序是稳定不变的，那么垂直柱就能学会这一顺序，它们就会以适当的序列接连兴奋起来。

第 1 层输入的另一半信息是来自体系中更高层级区域中的第 6 层，这些信息更稳定。如果你的垂直柱是音程的话，那么它就是曲子的名字；如果你的垂直柱是音素的话，那么它就是听到的单词；如果你的垂直柱是口语单词的话，那么从上面传来的信号就是你正在说的话。因此，第 1 层中的信息同时表征着序列的名字和序列中最后的那个项目。这样，一个特定

的垂直柱由于能被许多不同的序列分享而不会产生混淆。垂直柱学会了在适当的上下文中以正确的顺序产生兴奋。

在继续往下讨论之前，有一点我要指出来，那就是第 1 层中的突触，它们不是仅有的会预感垂直柱何时兴奋的突触。正如我前面所讲过的，细胞从许多邻近的垂直柱那里接受输入，也向它们发送信息。前面我们曾说过，有超过 90% 的突触是来自垂直柱以外的细胞，而且这些突触中的大部分都不是来自第 1 层的。比如，第 2 层、第 3 层和第 5 层中的细胞在第 1 层中有成千上万的突触，同时在自己的这个层次中也有成千上万的突触。这里全部的思想就是细胞需要信息以帮助它们预测自己何时会被下面的区域驱动。通常情况下，邻近垂直柱中的兴奋有着很强关联性，因此，我们能看到有许多伸展到邻近垂直柱的直接连接。比如，如果一条直线移过你的视野，那么它就会激活一系列连续的垂直柱。但是，用以预测一个垂直柱是否需要激活的信息则常常是更全局性的，这也是第 1 层的突触所要起的作用。如果你是一个细胞或一个垂直柱，你并不知道这些突触意味着什么，你所知道的仅仅是它们能帮助你预测自己何时需要兴奋起来。

现在让我们来看看皮层区域是如何为一个已了解的序列命名的。我们再一次来假设你是皮层中的一个区域。你的兴奋垂直柱会随着每一个新的输入而有所变化。你已经成功地学得了垂直柱被激活的顺序，也就是说，你的垂直柱中有一些细胞会先于来自下级区域的输入到达之前而兴奋起来。那么你发送给

皮层体系中更高区域的是什麼信息呢?我们已经知道,你的第2层和第3层细胞会将轴突延伸到较高的区域,这些细胞的兴奋状态就是向更高区域发送的输入。但这样会有一个问题。为了让整个体系能正常运作,你不得不在学习序列期间传递一个恒定的模式。你得传递序列的“名字”,而不是它的细节。在学习序列之前,你可以传递细节,但在你了解了序列之后,当你成功地预测哪些垂直柱会被激活时,你应该只传递恒定模式。然而,到目前为止,你还没有为序列命名。不论能否预测到变化着的模式,你都会将它们中的每一个进行传递。当一个垂直柱变得兴奋时,它的第2层和第3层细胞就会沿着体系向上发送一个新的信号。大脑皮层需要有一种机制,即在学习序列的过程中能让发送给下一区域的输入保持恒定。大脑皮层能使垂直柱在预测到自己要被激活时不让第2层和第3层的细胞有所输出,而当它预测不了时,仍让第2层和第3层的细胞保持兴奋。这是产生恒定模式的唯一方式。

现在我们还不甚了解大脑皮层,也无法说清楚它到底是如何做到这一点的。我认为有几种方法能让脑皮层做到这一点。我将介绍目前我认为最满意的观点——但请记住,比起谈到的具体方法来,更重要的是思想。产生恒定的“名字”模式是我们理论模型的要求。现在我所能说的只是这个命名过程存在着比较合理的机制。

现在我们来假设你是一个垂直柱,见图10所示的垂直柱。我们想知道,当你在能预测自己的兴奋状态时,是如何学会将恒定模式传递给较高的区域的,而当你不能预测自己兴奋状态时,又是如何将变化模式传递给较高区域的。我们先假定在第2

层和第3层有若干不同类别的细胞。(除了几类抑制细胞以外,许多解剖学家还将其区分为第3层a型和第3层b型这样两种细胞,因此,这个假设并不是没有道理的。)

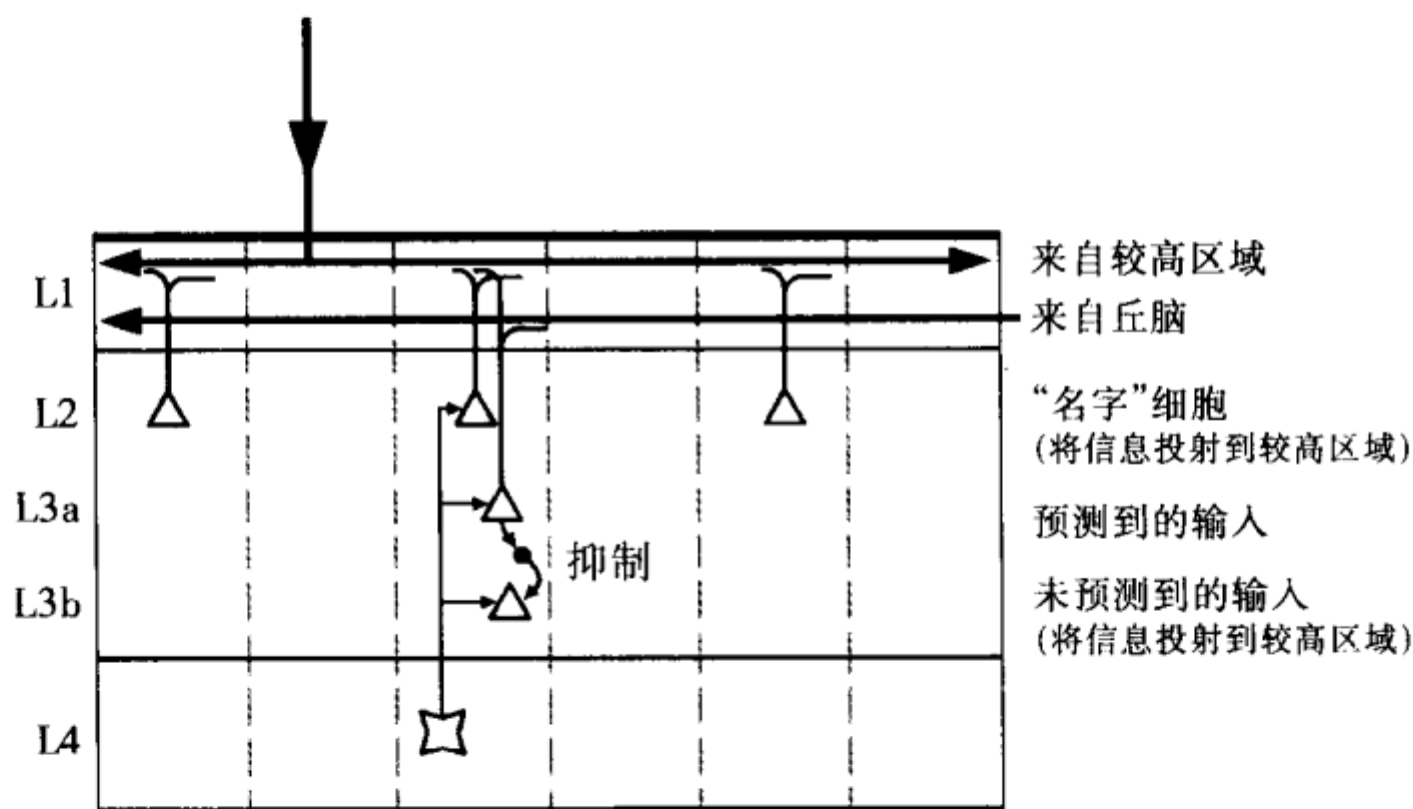


图 10 预测中抑制细胞的作用

我们还假定有一类细胞,我们把它叫做第2层细胞,它们在学习序列的过程中是一直保持兴奋状态的。作为一个整体,这些细胞表征着序列的“名字”。只要我们的皮层区域能预测接下来哪些垂直柱会兴奋起来,它们就会向较高的区域传递恒定的模式。假如我们的皮层区域有一个由5种不同模式组成的序列,那么,那些表征这5种模式的垂直柱中的第2层细胞,就会在我们处于这一序列当中时一直保持兴奋。它们就是这一序列的“名字”。

下面,我们再假设还有一种细胞,即第3层b型细胞,它们在垂直柱成功预测出输入时不产生兴奋,而当垂直柱不能预测兴奋状态时却能兴奋起来。那么第3层b型细胞就是表征预料之外模式的。当一个垂直柱出乎意料地变得兴奋时,这种细

胞就会兴奋；每当一个垂直柱在学习之前变得兴奋时，它也会兴奋。但当一个垂直柱学会了预测自己的兴奋状态时，第3层b型细胞就不会兴奋了。所以，第2层细胞和第3层b型细胞一起满足了我们的要求，学习之前，这两种细胞都会随垂直柱的状态而改变兴奋状态，但经过“训练”之后，第2层细胞就会恒定兴奋，而第3层b型细胞则一直保持安静。

这些细胞又是怎样学会这些的呢？首先，让我们来看看当垂直柱成功地预测到自己的兴奋状态时，第3层b型细胞是如何被抑制的。我们不妨认为，在第3层b型细胞的上方有另一类细胞，叫第3层a型细胞。这样的细胞也有树突延伸到第1层。它唯一的作用就是：当它在第1层中遇到合适的模式时，就会阻止第3层b型细胞产生兴奋；当第3层a型细胞在第1层中见到已学过的模式时，它就会迅速激活一种抑制细胞，以阻止第3层b型细胞的兴奋。这就是当垂直柱正确地预测了自己的兴奋状态时，它用以阻止第3层b型细胞兴奋的途径。

现在，我们再来看看更难的任务——在已知模式序列中使第2层细胞始终保持兴奋。这一任务更困难，其原因在于许多不同垂直柱中各类第2层细胞要同时保持兴奋，哪怕它们自己的垂直柱并没有被激活。我认为可以这样来解释这种现象：第2层细胞可以学会完全被更高层的区域所驱动，它们会优先与来自较高皮层区域的第6层细胞的轴突形成突触。这样，第2层细胞就表征着来自较高区域的恒定“名字”模式。当较高的皮层区域向下方区域的第1层发送一个模式时，下方区域的一组第2层细胞就会兴奋起来，它们代表着属于这一序列的所有垂直柱。因为这些第2层细胞也将信息投射回较高的皮层区

域，所以这些细胞也就成了一组半稳定的状态。(这些细胞并不见得会恒定地保持兴奋，它们大概只是有节奏地兴奋。)这个过程就好像是较高区域向下方区域的第 1 层发送了一个曲调的名字，使一组第 2 层细胞兴奋，每一个第 2 层细胞都对应着听到曲调时会产生兴奋的垂直柱。

所有这些机制合起来，就能让大脑皮层学习序列、做出预测以及为序列形成恒定表征(或“命名”)。这些是形成恒定表征的基本操作。

那么，我们又是如何对从未见过的事件进行预测的呢？我们是如何在一个输入的众多解释中进行取舍的呢？皮层区域又是如何基于恒定记忆而做出具体预测的呢？关于这个问题，之前我举过几个例子。比如，当你的记忆仅仅调用了两个音符间的音程时，你能预测到曲调中的下一个音符；还有关于火车的例子和背诵葛底斯堡演讲的例子。在所举的这些例子当中，解决问题的唯一途径就是利用最后的特定信息将恒定预测转化为具体预测。换言之，从大脑皮层的角度来说，我们必须将前馈信息(实际的输入)与反馈信息(恒定形式的预测)相结合。

下面这个例子可以说明我对此是怎样认为的。比如说，你的皮层区域被告知要期待一个 5 度音程。区域中的垂直柱表征着所有可能的具体音程，比如 C - G、D - A、C - E 等，你必须决定需要激活哪个垂直柱。当上方区域告诉你要期待一个 5 度音程时，它就激活了所有表征 5 度音程(比如，C - G、D - A 以及 E - B 等)的垂直柱中的第 2 层细胞，而表征其他音程的垂直

柱中的第 2 层细胞将不被激活。现在，你必须在所有可能表征 5 度音程的垂直柱中选择出一个来。皮层区域的输入是具体的音符，如果你最后听到的音符是 D，那么所有表征与 D 有关的音程的垂直柱，比如，D - E 和 D - B，就获得了部分的输入。因此，现在在第 2 层中的所有表征 5 度音程的垂直柱都兴奋起来了，而在第 4 层中，所有表征与 D 有关的音程的垂直柱也获得了部分信息。这两组垂直柱的交集就是我们的答案，也就是表征 D - A 5 度音程的垂直柱(详见图 11)。

大脑皮层又是如何找到这个交集的呢?记得在前面我提到

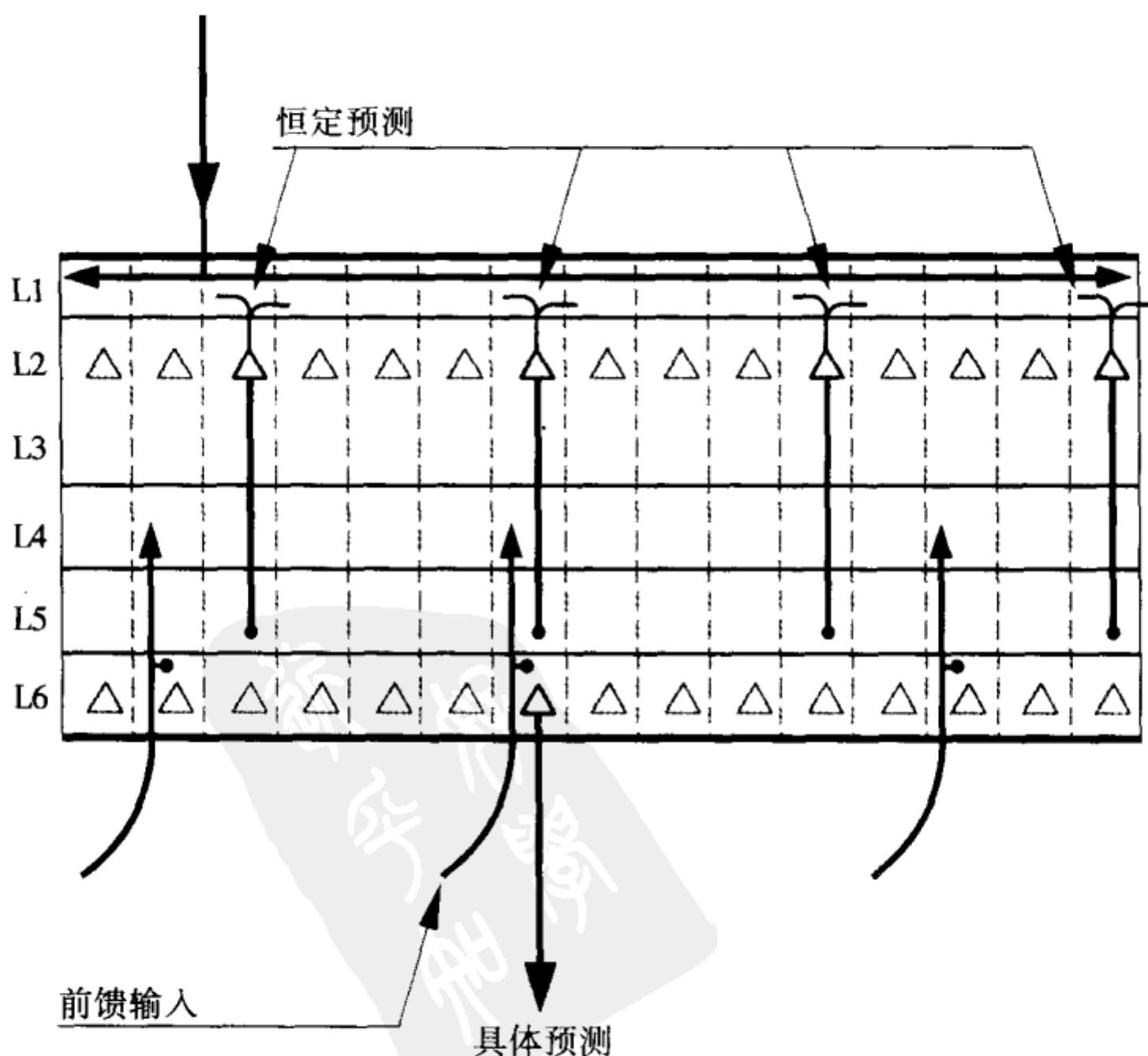


图 11

过，第2层和第3层的细胞在离开皮层时通常会在第5层形成突触，同样地，靠近第4层的来自下面区域的轴突在第6层会形成突触。这两种突触（一个自上而下，一个自下而上）正是我们需要的。接受了这两种兴奋输入的第6层细胞就会兴奋。第6层细胞表征着皮层区域认为正在发生的状况，也就是一个具体的预测。如果第6层细胞会说话，它就会说：“我属于一个表征某事物的垂直柱。现在的情形是，我所属的垂直柱表征着D-A音程，其他垂直柱则表征其他的事物。我代表着我的皮层区域。一旦我兴奋起来，就意味着我们认为D-A要么是已经出现，要么就是即将出现。我之所以兴奋是因为来自双耳的自下而上的输入使第4层细胞激活了整个垂直柱。或者我的兴奋状态可能意味着我们识别了一个曲子，并正在预测下一个具体的音程。不管是哪种情况，我的任务就是告诉下面的皮层区域我们认为发生了什么。我代表着我们对世界的解读，不管它是现实的还是想像的。”

我再用另一个心理图画来描绘这一机制。设想一下，我们有两张钻了许多小孔的纸片，其中一张纸片上的小孔代表着那些第2层或第3层兴奋的细胞垂直柱，也就是我们的恒定表征。另一张纸片上的小孔代表着来自下方区域的部分输入。如果将两张纸片重叠的话，你会发现，其中一些小孔能对齐，而另外一些则对不齐。对齐了的小孔就代表着我们认为需要激活的垂直柱。

这一机制不但能产生具体预测，还能消除来自感觉输入的歧义。有时候，一个皮层区域的输入可能是模棱两可的，就好像我们所提过的关于彩色纸片的例子，或者听到了一个模棱两

可的单词。这种自下而上、自上而下的匹配机制，能让我们在两个或更多的解读中做出取舍。一旦做出了选择，你就会将你所选的解读传递给下面的区域。在你的一生中，只要你是清醒的，你的新皮层中每一个区域时时刻刻都在将由上方区域激活的一组期待垂直柱与由下方区域激活的一组实际观测垂直柱相比较。这两组的交集就是我们所感知的。如果我们的输入是完美的，且我们的预测也是完美的，那么，那组感知垂直柱就总会出现在那组预测垂直柱中。但并不会总是这么一致。这种结合部分预测与部分输入的方法能消除有歧义的输入，能填补缺失的信息，也能在不同的解读中做出选择。我们就是这样通过将—个期待的音高——恒定的音程与最后听到的音符相结合，来预测曲子中接下来将要听到的具体音符的；我们也是这样来判断一幅画是一个花瓶还是两张脸的；同样，我们也是用这种方法将运动指令流一分为二，分别写出或说出葛底斯堡演讲的。

最后，第6层细胞除了把信息投射到较低的皮层区域外，它们还将输出信息发送回自己所属垂直柱的第4层细胞，这样一来，我们的预测就成了输入。想入非非或思考问题时，我们就是这么做的，这能让我们知道自己预测的结果。每天我们都花费一些时间在做这样的事情——规划未来、排演讲语以及担心将要发生的事。斯蒂芬·格罗斯伯格，一位长期从事大脑皮层研究的科学家，把这种情况叫做“折叠反馈(folded feedback)”。而我更喜欢“想像(imagining)”这一提法。

在结束这一部分之前，我们再来看看最后一个问题。我曾

有好几次都指出，我们所看到、听到和触摸到的结果往往取决于我们自己的行动。看什么，取决于我们的眼睛在扫视何处、我们的头又是如何转动的。触摸什么，取决于我们怎样移动自己的手臂和手指。会听到什么，有时取决于我们的所说和所做。

因此，要想预测自己下一步的感受，我们就得知晓自己目前采取的是什么样的行动。运动行为与感官知觉相互是高度依赖的。如果我们下一步的感觉每每取决于我们自己的行动，那么我们如何进行预测呢？所幸的是，对这一问题我有一个漂亮而又令人吃惊的答案，虽说其中有许多细节尚不清楚。

第一个让人感到吃惊的发现就是，知觉和行为几乎就是同一的。正如我前面说过的，大脑皮层中如果不是全部的话，也是大部分区域，甚至包括视觉区都参与了运动的产生过程。投射到丘脑和第1层的第5层细胞似乎也有运动功能，因为它们同时也将信息投射到古脑的运动区。这样，对于“刚刚发生了什么”——包括感觉和运动两方面的知识，第1层都能获知。

令人感到惊奇的第二点，同时也是第一点的结果，那就是，运动行为也必须由恒定表征体系来表征。通过以恒定方式来周详地考虑将要做的事，你就会产生完成特定活动所需的动作。运动指令在沿体系向下传播的过程中被转化为完成预期活动所需的复杂而细节化的序列，这在“运动”皮层区和“感觉”皮层区都有发生，从而模糊了彼此的差别。如果视觉皮层的IT区感知到了“鼻子”，在向“眼睛”的表征切换时，就产生了让预测（由看到“鼻子”而预测到能看到“眼睛”——译者注）变成现实所需的一次扫视。由看“鼻子”转换成看“眼睛”所需的特定扫视，会根据所看的面孔在何处而有所不同：较近

的面孔需要较大幅度的扫视；面孔距离较远时，则只需要较小幅度的扫视就行。侧着的面孔与正面的面孔所需的扫视的角度也是不同的。随着看“眼睛”的预测向 V1 区流动，所需扫视的细节也随之确定下来。预测向下走得越低，扫视就变得越是具体，最后，经过扫视，你的视网膜凹正好落在要观察的对象上，或者很接近那个对象的地方。

我们再来看看另一个例子。我的身体要从起居室走到厨房，这时，大脑的任务就是要在心理上将起居室的恒定表征转换到厨房的恒定表征，这样的切换导致了序列展开的复杂过程。在从起居室走到厨房的过程中，产生了关于我将所看、所触和所听的预测序列，同时也产生了能使我从起居室走到厨房的运动指令，并导致在此过程中我眼睛的移动。模式在皮层体系中上下流动的过程中，预测和运动行为彼此很好地合作着。只要你的行为有所涉及，你的预测就不但要先于感觉，而且还会决定感觉。这听起来让人感到有点奇怪。对序列中下一模式的思考，会引起对你接下来将要体验到什么的预测，即所谓阶流式预测(cascading prediction)。随着阶流式预测的展开，就产生了完成预测所需的运动指令。思考、预测和行动都属于沿皮层体系下行序列的同一个展开过程。

通过思考而“行动”，以及知觉和运动行为的平行展开过程，是所谓面向目标的行为的核心内容。面向目标的行为是机器人研究的神圣目标，它被置入皮层的结构之中。

当然，我们可以关掉运动行为。我可以想像看到了某个事物而不必真正看见它，我也可以想像已走到厨房而不必真正那样去做。但是，“思考做某事”确实是我们完成这件事的开始。

上行与下行

让我们先来回想一下，信息是如何在皮层体系内上下流动的。当你在世界中走动时，变化着的输入涌进大脑皮层较低的区域。每一区域都会试图把输入流解读为属于已知模式序列的一部分。垂直柱试图预测自己的兴奋状态。如果垂直柱预测到了自己的兴奋状态，它们就向上方区域传递一个稳定的模式，或者序列的名字。就好像这个区域会说：“我听到了一首歌，这是歌的名字，我能处理有关它的细节。”

但如果一个预料外的模式到达了，比如，听到了一个预料外的音符，那又会怎样呢？或者说，假如我们看到的并不是面孔的一部分，那又会怎样呢？预料外的模式会被自动传递到较高的皮层区域。当不属于期待序列的第3层b型细胞兴奋时，这种情况自然而然就发生了。较高的区域也许能将这一新的模式解释成自己已知序列的下一部分。它也许会说：“噢，我看到了一个新的音符。也许这是歌曲集中下一首歌的第一个音符，好像就是这样。因此，我预测现在我们已经转到下一首歌了。喂，下方的区域，这就是下面这首歌的名字，我想你听的就是它了。”但是，如果连这样的识别过程都没有发生，这个预料外的模式就会一直沿着皮层体系往上传递，直到某个区域能把它解读成自己正常的事件序列。预料外模式传递得越高，就有越多的皮层区域参与到消除意外输入的处理过程中来。最后，当体系的上方有一个区域认为自己能解读这个预料外的事件时，它就会产生一个新的预测，这个新的预测就沿着体系尽可能地向下传播。如果这个新的预测还是不对，就会有错误被

检测到，这样，这个预料外模式又将再一次沿体系往上传递，直到有某个区域能把它识别为自己兴奋序列的一部分。这样，我们看到观测模式沿体系向上传递，而预测则沿体系向下传播。在理想状态下，如果一个世界是已知的并且是可预测的，那么，大部分的上下模式流主要是快速地发生在皮层的较低区域。大脑迅速地尝试着找到自己世界模型中与预料外输入相一致的部分。只有在这个时候，大脑才能有理由知道接下来要期待什么。

假设我走进自己家的一间熟悉的房间，就不可能有任何错误会沿大脑皮层传播。关于我家的已熟练掌握的序列在视觉和运动体系的低级区域就能很好地被处理。我非常熟悉自己的房间，甚至可以在黑暗中摸索着四处走动。因为我对周围环境很熟悉，所以大脑皮层的大部分区域能得到有效的释放，可以用来完成其他的一些任务，比如思考关于大脑的问题和写书等等。然而，如果我是在一间不熟悉的房间，尤其是在一间与我以前所见过的很不一样的房间里，我不但要四处打量以确定自己的位置，而且预料外的模式会不断传到皮层体系中高层区域。感觉体验与已学序列之间越是不匹配，就会有越多的错误向上传递。在这样的新环境中，我不可能再思考有关大脑的问题，因为大部分皮层区域都在忙于处理漫步房间的事宜。这是那些步下飞机踏上异国土地的人们都曾有过体验：即使异国的道路与你熟悉的道路大同小异，但汽车可能会奔驰于马路的另一边；货币也很陌生，语言也不通，找厕所也需要花费所有的大脑资源。初到异国他乡，走在路上，我们是无暇排演讲话的。

顿悟的感觉，“啊哈！”——突然明白后所发出的感叹，



图 12 你看到了斑点狗吗

都能用这个模型来解释。想像一下，假设你正在看一幅模糊的图片。画面上满是墨水团和七零八散的线条，它好像什么都不是，丝毫没有意义。当大脑皮层没有找到与输入相匹配的记忆时，就产生了混乱。你的眼睛细细审视画面的每个角落，新的输入直接传到皮层体系的高层区域，高层的皮层区域尝试给出许多不同的假设，但每当这些预测沿体系向下传播的时候，它们都与输入发生了冲突，皮层不得不再次尝试新的假设。在整个混乱的过程中，你的大脑完全投入到了理解图画的任务之中。终于，你的高层皮层区域做出了正确的预测。这时候，这个预测从皮层体系的最高层出发，成功地通过了一个又一个的

区域，最后到达了皮层的底层。在不到 1 秒钟的时间内，大脑皮层的每一区域都得到了与输入相匹配的序列，不再有错误传到高层。你知道自己已经看明白了这幅图画，看到了一只走在点线之间的斑点狗(见图 12)。

反馈真能完成这一任务吗？

几十年来，我们一直都认为皮层体系中的连接是相互的，如果区域 A 将信息投射到区域 B，那么区域 B 也会把反馈信息投射回区域 A。一般情况下，反馈轴突纤维比前馈轴突纤维要多。然而，即使这一看法被人们广泛接受，人们却普遍认为反馈在大脑中的作用很小，仅仅起“调节”的作用。那种认为反馈信号能及时且准确地引起第 2 层中多组细胞兴奋的看法，并没有得到广大神经科学家的认同。

为什么会这样呢？其中部分的原因在前面已有所提及，那就是，如果你不接受预测的关键作用，也就没有必要虑及反馈了。如果你认为信息直接流向运动体系，那为什么还需要反馈呢？不考虑反馈的另一个原因就是，反馈信号遍及第 1 层的广大区域。我们一般认为大面积分布的信号对许多神经元只能起到很小的作用，大脑中确实存在着一些这样的调节信号，它们不针对特定的神经元产生作用，而只是改变诸如警觉等的全局特征。

忽视反馈作用的最后一个原因在于，没有多少科学家认为单个神经元是独立工作的。典型的神经元有成千上万的突触，这些突触有些远离细胞体，另一些则离细胞体很近。靠近细胞体的突触对细胞的兴奋有很强的影响。细胞体附近如果有十多

个兴奋的突触，那么它就可能释放电脉冲。这是大家都知道的。然而，绝大多数突触并不靠近细胞体，它们分布在有着大脑一样结构的细胞树突上的广大区域。正因为这些突触远离细胞体，所以科学家大多都认为，到达其中一个突触的电脉冲对其所在神经元所产生的作用很弱，甚至难以察觉。而对那些更远的突触，电脉冲可能在还没到达细胞体之前就已经消散了。

沿皮层体系向上传递的信息是通过靠近细胞体的突触来传递的，这被看作是一条基本原则。因此，我们可以更加肯定，沿体系向上传播的信息会流经一个个的区域，同时，沿体系向下传播的反馈信息是通过远离细胞体的突触而传递的，这也被看作是一条基本规则。第2层、第3层和第5层中的细胞都将树突伸展到第1层，并在第1层形成许多突触。第1层遍布着突触，但这些突触都远离位于第2层、第3层和第5层的细胞体。此外，第2层中任何一个特定的细胞与任一特定的反馈纤维只会形成几个突触(如果有的话)。因此，一些科学家对第1层的短暂模式会精确地引起第2层、第3层和第5层中的一组细胞兴奋这一看法是持反对意见的。但我所提理论恰恰要求这样。

解决这一难题，关键是要认识到神经元并不是以传统模型中的方式在工作。实际上，近年来有越来越多的科学家提出，远距离的纤细树突上的突触对细胞兴奋有着很积极和明确的作用。在这些模型里，这些远距离的突触与靠近细胞体的粗厚树突上的突触有着不同的表现。比如，如果纤细树突上有两个彼此很接近的突触，那么它们就会充当“共同探测器(coincidence detector)”的角色。也就是说，如果在一个很小的时间窗内，这两个突触都收到了一个尖脉冲输入，那么即使它

们都远离了自己的细胞体，也会对所属细胞产生很大的影响——它们会使细胞体也产生一个尖脉冲。神经元的树突到底是如何工作的，这仍然是个谜，所以，在此我不会多谈。重要的是，大脑皮层的记忆-预测模型要求远离细胞体的突触能探测到特定的模式。

回头再看，认为一个神经元上成千上万的突触只起调节作用的想法似乎是很愚蠢的。大量的反馈信息和大量的突触之所以存在是自有道理的。根据这样的看法，我们可以说，当典型的神经元在纤细的树突上形成突触时，它们就能够学习反馈纤维上的上百个精确耦合。这意味着，从哪个反馈模式会使其激活这个角度来考虑，我们新皮层的每一个垂直柱都具有高度的灵活性。同时也意味着，任一特定的特征都能与成千上万个不同的对象和序列产生精确的联系。我的模型要求反馈要迅速且精确，即细胞必须在它们发现远离自己的树突上一旦有精确的耦合时就产生兴奋。这一点在那些新的神经元模型中已被考虑到了。

大脑皮层如何学习

皮层的所有层次中任何细胞都有突触，而且大部分突触都可以通过经验进行修改。我们不妨可以说，学习和记忆存在于所有的层次和所有的垂直柱中，以及所有的皮层区域中。

在本书的前面，我提到过以加拿大神经心理学家唐纳德·海布(Donald O. Hebb)的名字而命名的海布学习法则。其要点很简单：如果两个神经元同时兴奋的话，它们之间的突触就会得到加强。现在我们知道海布基本上是正确的。当然，世上万物都不会是这样简单的，人脑中的许多细节就更加复杂了。我

们的神经系统有许多海布学习法则的变体，例如，一些突触会根据神经信号时长的微小变化而改变其强度，一些突触变化很短暂，而另一些则较长。但海布仅仅是为关于皮层学习的研究提供了一个框架而已，它不是一个最终的理论，但这一框架是相当有用的。

海布学习法可以解释我在本章中提到的大部分皮层行为。请记住，还是在 20 世纪 70 年代就已经证明了，利用经典海布学习算法的自-联想记忆系统就可以学习空间模式和模式的序列。它的主要问题在于，这样的记忆不能很好地应对变化。根据本书提出的理论，大脑皮层能突破这方面的限制，其途径部分是通过将自-联想记忆纳入皮层体系之中，部分是通过复杂的垂直柱结构。本章主要讨论的就是皮层体系及其工作原理，因为体系结构是大脑皮层之所以强大的原因。因此，我并不是要详细讨论每一个细胞是如何学这学那的，我只想论及一些皮层体系学习的基本原则。

在你出生之际，你的大脑皮层对世界基本上是一无所知。它不了解你的语言，你所处的文化，你的家，你的城市，歌谣以及将伴随你成长的人，总之是一无所知。所有这些关于世界结构的信息，都要通过学习来获得。学习的两个基本要素就是形成对模式的分类和建立序列，这两个相辅相成的记忆组件相互作用着。当一个皮层区域学习序列的时候，到达第 4 层细胞的输入就在变化。因此，这些第 4 层细胞就会学着进行新的分类，并改变投射回第 1 层的模式，进而对序列产生影响。

形成序列的基本过程就是将属于同一事物的模式归为一个组。有一种方法就是将时间上连续发生的模式归为一组。如果

一个孩子手拿玩具并慢慢移动它，她的大脑会认定她的视网膜上对于该玩具的影像是不同时刻下同一物体的影像，因此，这一组变化着的模式可以联系到一起。而在另一些场合下，你却需要借助外部指令来决定哪些模式是一个整体。要想了解苹果和香蕉是水果，而坚果和芹菜却不是，就需要老师指导你将前两者归为水果。不管是哪种方式，你的大脑都在慢慢建立起归为一组的模式序列。但是，当一个皮层区域在建立序列时，下一个区域的输入就在改变，输入从主要表征单个的模式变成了表征一组一组的模式，一个区域的输入就从音符变成了曲调，从字母变成了单词，从鼻子变成了面孔，等等。因为到达一个区域的自下而上的输入变得更加“面向对象”，所以皮层的较高区域现在可以学习这些更高层级对象的序列了。一个曾形成字母序列的区域，现在形成了单词序列。这一学习过程给人带来了意想不到的结果：在反复的学习过程中，对象的表征会沿着皮层体系向下移动。在你生命的前几年里，你对世界的记忆首先形成在皮层的较高区域，但随着你的进一步学习，这些记忆会在皮层体系的越来越低的区域中重新形成。这并不是说大脑将它们向下转移，而是指它不得不一次又一次地重新学习。（我并不是说，所有的记忆都是从皮层的顶层开始的。当然，记忆的实际形成过程更为复杂。我相信第4层的模式分类是先从底部开始，然后往上移动的。但在此过程中，我们开始形成序列，随后这些序列往下移动。我指的是对模式的记忆会在皮层的越来越低的区域中重新形成。）随着简单的表征往下移，位于皮层底部的区域就能学习更复杂、更细微的模式了。

你可以通过观察孩子的学习过程，来了解分层记忆是如何

形成并向下移动的。想一想我们是怎样学习、阅读的吧。我们首先学会的是识别单个的印刷字母，这是一个缓慢而艰难的任务，它要求有意识地去努力。然后我们要识记简单的词汇，这一过程开始也是很困难且缓慢的，即使是学习最简单的词(三字母词)也是一样。孩子可以挨个念出每一个字母，但需要大量的练习才能把它们认作是一个完整的单词。学习了简单词汇以后，我们开始艰难地学习多音节单词。起先，我们会念出每一个音节，然后把它们连起来，就像我们学简单词汇时把字母连起来一样。经过多年的练习，我们才能读得很快。再往后我们会达到这样一种程度，不再看每一个单个的字母，而是一眼就能识别整个的单词或短语。这并不是因为我们读得快的缘故，实际上，这是因为我们把单词和短语当作一个整体来识别了。那么，当我们识别一个完整单词的时候，我们还能看到字母吗？答案既是肯定的，也是否定的。视网膜看到的是字母，因此 V1 区看到的也是字母。但识别字母只发生在皮层体系的较低层次，比如说 V2 区或 V4 区。当信号到达 IT 区时，表征的不再是字母了。当诸如字母等简单对象的记忆在体系中往下移动时，皮层的较高区域就能够学习单词和短语等更复杂的对象了。

另一个例子是学识乐谱。首先你得全神贯注于每一个音符。通过练习，你开始识别常见的音符序列，然后是整个乐句。经过大量练习以后，你就好像根本没有看见大部分的音符一样。活页乐谱放在那儿仅仅是帮你回忆乐曲的主要结构，细节性的序列已经储存在更低的层次中了。这类学习在运动和感觉区域都是存在的。

孩子的大脑在识别输入和做出预测方面都是比较慢的，这是因为服务于这些任务的记忆储存在皮层体系的较高层级中。为了解决冲突，信息不得不上下贯穿整个皮层，甚至还需要经过多次的反复。神经信号要上下贯穿整个皮层体系是需要时间的。孩子大脑的顶部还没有形成复杂的序列，因此还不能识别和回放复杂模式，还理解不了世界的高级结构。与成人相比，孩子的语言、音乐以及社会关系都很简单。

如果你反复研究某一类事物，你的大脑皮层将在较低的层级中形成对它们的记忆表征，这就把皮层的较高层级给释放出来了，可以利用它来学习更细微、更复杂的关系。根据这一理论，专家就是这样被培养出来的。

在设计电脑的工作当中，我在看到产品时就能很快发现其设计上的瑕疵，很多人对此感到惊讶不已。经历了 25 年的电脑设计工作，在与移动计算设备有关的问题上，我形成了一个比普通要强得多的模型。同样地，一位有经验的父母是很容易就能知道孩子为什么会不安的，但刚做父母的人处理这样的情形来则可能要费更多的周折。有经验的经理人能很容易发现团队中结构上的问题和优势，而对新手来说，他们则很难看到这些。他们都有着同样的输入，但新手的模型显然不够复杂，缺乏高级。在所有这些以及更多的事例当中，我们都是从学习最基本的结构和基础知识开始的。随着时间的推移，我们会将这些知识下移到皮层体系的下方层级中，这样，皮层的上方层级就有机会学习更高级的结构了。正是这些更高级的结构，使我们变得很有经验。专家和天才有着比普通人更能观察结构之结构、模式之模式的大脑。你也能通过练习而成为专家，但这其

中当然也还包含遗传方面的因素。

海马：处于最高层

在新皮层下有相互能沟通的三大脑结构，它们是基底神经节、小脑和海马。这三大结构都先于新皮层而存在。粗略地说，基底神经节是原始的运动系统，小脑主要学习事件间精确的时间关系，海马则储存具体事件及与地点有关的记忆。从某种意义上讲，新皮层包含有它们原有的功能。比如，一个生来就小脑缺损的人在时间感觉方面会出问题，因此在行动的时候需要刻意地去注意，除此以外，在其他方面他都很正常。

我们知道，新皮层负责所有的运动序列，它能直接控制你四肢的活动。我们并不是说基底神经节就不重要了，我们只是说新皮层已接管了大部分的运动控制。因此，我是在没有提及基底神经节和小脑的情形下来描述新皮层的全部功能的。有些科学家可能不赞同这一假设，但我在本书中和实践中都是这样假定的。

然而，海马却与众不同。我们对大脑中这一区域的研究最为深入，因为它对形成新的记忆是至关重要的。如果你失去了左、右两半海马（像许多神经系统的组成部分一样，海马在大脑中也为左、右两半），你就失去了形成新记忆的能力。没有了海马，你可以照常说话、走路、看东西和听东西，乍一看你几乎就是个正常人。但实际上，你受了很大的损伤：你记不住所有新的事物；你能记起失去海马前许多年认识的朋友，但你却记不住一个新遇见的人；即使在一年里，你每天去看你的私人医生 5 次，但你每次见到他时都会觉得是第一次见到他。你对

失去海马以后发生的一切事情都无法记住。

多年来，我一直不愿意讨论海马，因为它对我来说没有任何意义。很显然，它对学习相当重要，但我们所掌握的大部分知识最后却不是储存在这里。而新皮层才是信息存放之处。人们对海马的传统认识是：新的记忆先在那里形成，然后，几天、几个星期或者几个月之后，这些新的记忆才被转移到新皮层中。我根本就不这么认为。我们知道，景物、声音、所触之物——一切感觉数据流，都直接流入皮层的感觉区，而不是先经过海马。在我看来，这些感觉信息应该是在皮层中自动形成新的记忆。那为什么大脑皮层学习的时候还需要海马呢？像海马这样一个独立的结构又是怎样干预或阻止皮层中的学习的呢？它又是怎样在若干时间之后再与信息转移回皮层的呢？

我决定先把海马放到一边，心想总有一天它的作用会明朗起来。这一天终于在 2002 年末出现了，那正是我着手写这本书的时候。布鲁诺·奥斯豪森(Bruno Olshausen)，我在红杉神经科学研究中心的一名同事指出，海马与新皮层之间的连接表明海马是新皮层的最高区域，而不是一个独立的结构。这样看来，海马应该处于新皮层金字塔的顶端，也就是图 5 中的最上面的部分。从进化的角度来看，新皮层似乎是处于海马和其他大脑组织之间。显然，认为海马是皮层体系的最顶端的看法也有些时日了，对此我却不甚了解。我与几位研究海马的专家谈过，我要他们解释一下这个外形酷似海马的结构是如何将记忆转移到皮层的。他们谁都没法解释。并且他们中谁都没有指出海马是大脑皮层金字塔体系的顶点，大概这是因为在生理上，海马不仅位于皮层金字塔的顶端，而且还与许多大脑的较古老

部分直接相连吧。

但是，我马上意识到，这一观点正是解决我当前困惑的关键。

想一想从眼睛、耳朵和皮肤传到新皮层的信息。新皮层的每一个区域都试图理解这些信息的意义，每个区域都试图将输入理解为自己已知的序列。如果确实理解了输入，它就会说：

“这个我明白。它是我已经看到的事物的一部分。我不会传递细节的信息。”如果一个区域理解不了当前的输入，它就会把信息往上传递，直到有更高的区域能理解它。但是，一个完全崭新的模式会逐步沿着皮层体系上升到越来越高的地方。每一个更高层次的区域都会说：“我不知道这是什么。对此我没有预期，上面的伙计，你何不看看这是什么呢？”最后的结果是，当信息到达皮层体系顶部的时候，剩下的只是根据以前的经验所无法理解的信息了。现在你得到的就是崭新的、无法理解的那部分输入。

每天我们都会遇到许多新的事物，并且这些新事物都会直接到达皮层的顶端。比如说，报纸上的一条新闻，早上结识的那个人的名字，或者在回家路上看到的交通事故，等等。这些无法解释也没有预期到的事物的新记忆进入到海马，并储存在那里。但这些信息不会一直存在那里。这些信息要么被转移到皮层中，要么就永久地丢失了。

我注意到这样一种现象，随着年岁的增长，我感到要记住新事物越来越困难了。比如，我的孩子能记住去年看过的大部分电影中的细节，而我就想不起来。或许这是因为我所看过的多部影片对我而言毫无新鲜感，新电影的内容总是与看过的旧

影片的记忆有雷同之处，因此，这些新的信息不会到达我的海马。而对我的孩子而言，每一部电影的内容都比在我看来要更新一些，所以这些信息就会传到海马。如果这是事实的话，我们就可以说，你知道得越多，你记住的就会越少。

与新皮层有所不同的是，海马是一个非均质的结构，它是由若干专门化的区域组成的。它所擅长的就是，快速记住它所看到的任何模式。海马处于皮层金字塔的顶端，这是一个最佳的位置，能记住新的事物。这一位置对于调用这些新的记忆来说也是最佳的，它能让这些记忆转存到皮层体系中。皮层体系存储记忆是一个稍许缓慢的过程，而在海马里，你能快速记住一件新事物。你必须再三体验同一件事物，不管你是在实际体验还是在想像，它才会永久地存于皮层中。

上溯体系的第二通路

大脑皮层中还有第二个主要通道能使信息在区域间传递，通过这一通道，信息还能在体系中向上传递。这条第二通路是从有轴突伸入丘脑（这部分丘脑不同于我们先前讨论过的那部分）的第5层细胞开始，然后经由丘脑一直延伸到皮层中紧接着的较高区域。每当皮层的两个区域上下彼此直接相连时，它们也通过丘脑间接地相连。第二通路只沿体系向上传递信息，而不向下传递信息。因此，当我们沿皮层体系上溯的话，就会有两条通路：一条是连接上、下两个区域的直接通路，另一条则是经由丘脑的间接通路。

第二通路存在由丘脑细胞控制的两种工作状态。在第一种状态下，通路基本上是关闭的，因此信息不能经过。在第二种

状态下，信息在两个区域间精确地流动。有两位科学家描述过这第二条通路，并认为它可能与本章曾重点讨论过的直接通路一样重要（也许还更重要）。其中一位是纽约州立大学石溪分校的默里·舒曼（Murray Sherman），另一位是威斯康星大学医学院的雷·古勒雷（Ray Guillery）。

我对第二通路的作用是这样考虑的。

请读一下这个词：**imagination**。多数人只需瞥上一眼，凭一次注视就能认出这个词。现在再看一下单词中间的字母 i。再看一下 i 上的那个点。你的眼睛可以只注视同一个地方，但第一次你看见的是整个单词，第二次看见的是字母 i，而最后一次看到的是 i 上的点。请你注视着字母 i，同时尝试着在单词、字母和点之间切换你的感知。如果这样做有困难的话，请你就在注视着点的时候反复念出“点”、“i”和“imagination”。在每一种情况下，都有同样的信息进入到 V1 区，但当它到达 IT 区等较高的区域时，你感知的是不同的对象，它们有着不同的细节信息。IT 区知道如何去识别所有这三个对象。它能单独认出字母 i 上的点，也可以一眼就认出整个单词。但是，当你感知到整个单词的时候，V4 区、V2 区和 V1 区都在处理细节，IT 区知道的仅仅是单词本身。在阅读的时候，你通常是感知不到单个字母的，你感知到的是整个的单词和短语。但如果你愿意的话，也可以感知到单个字母。我们时时刻刻都在做这样的注意力切换，只是在一般情况下我们没意识到而已。在听背景音乐时，我可以一点儿也注意不到它的旋律，但如果我尝试一下，就可以分离出歌手的声音和贝司、吉他的声音来。进入到我头脑中的都是同样的声音，但我可以把注意力集中到某一种

感知上。每次挠脑袋时，手指的运动都会在脑袋里产生很大的声音，但通常你却感觉不到这个噪音。而如果你将注意力集中到它的上面，就可以清楚地听到声音了。这个例子又说明了感觉的输入通常是在皮层体系的较低区域被加工处理的，但如果你需要注意到它的话，它也可以被传送到高级区域。

我猜测，连接丘脑的第二通路就是我们去注意平时不会注意到的细节的机制。它绕过第2层对序列的分类，直接将原始信息送到皮层的较高区域。生物学家指出，第2条通路可以通过下面两种途径开启。其一是通过皮层中较高区域自己发出信号。这是当我要你注意你平常不会注意到的细节的时候你所采用的方法，比如要你看字母 i 上的点或者听挠头时候的声音。开启这条通路的第二种方法就是从下方传来的很强的预料外的信号。如果到达这条通路的输入足够强，它就会向较高区域发送一个唤醒信号，这样也可以开启通路。比如，如果我给你看一副面孔，然后问你它是什么，你会说“脸”。如果我还是给你看同一张脸，但这次鼻子上有一个奇怪的印记，你会先认出脸来，但很快视觉区的较低层次会注意到有什么不对劲的地方。这时，错误强制性地开启了这一通路。现在，细节信息可以取道第二通路，绕过正常情况下要进行的序列分类，你的注意力也就随之转移到了那个印记上。现在，你注意的将是那个印记而不是那副面孔。如果这个印记相当古怪，那么它将占据你全部的注意力。这样，不常见的事件能很快吸引你的注意力。这就是为什么我们无法避免将注意力集中到畸形事物和其他不常见模式上的原因。你的大脑会自动地完成这一切。然而，有时错误信息不够强，不足以开启第二通路。这就是为什

么在阅读的时候一个单词拼错了而我们却注意不到的原因。

结束语

发现并建立一个崭新的科学理论框架，要求我们寻找最简单的原理。这个简单的原理必须能整合并解释大量看似毫无联系的事实。这一过程带来的必然后果就是：我们可能会将问题过于简单化。这样，很多重要的细节可能会被忽视，事实也可能被曲解。如果这个理论框架有一定的道理，我们就肯定能发现其需要改进或修正的地方，比如或许我们走得太远了，或许走得还不够远，或许它根本就是错误的。

在这一章中，我提出了许多有关新皮层工作原理的设想，我很希望其中有一部分思想被证明是错误的，也许所有的想法都需要进行修订。我还有许多细节没有提到。大脑过于复杂，如果是神经科学家读这本书的话，他们就会觉得我在此仅仅是提供了有关人脑复杂性的一个很粗糙的描述。但我相信，这一框架整体上是站得住脚的。我希望，随着新的数据和理解的出现，细节的内容将会得到修正，但核心思想却能得以保留。

最后，你可能会觉得简单而庞大的记忆系统能产生所有人类行为这一思想难以接受。你我难道真是一个多级记忆系统吗？我们的生活、信念和雄心抱负都是储存在上万亿个细小的突触之上的吗？1984年，我开始专业从事编写计算机程序。以前我曾编过一些小型程序，而这一次却是我第一次编写一个带有图形用户界面的计算机程序，也是我第一次编写大型而复杂应用软件。我在为由 Grid Systems 公司创建的操作系统编写软件。因为有了视窗，多种字体，还有菜单，Grid 操作系统在当时算是

相当进步的。

有一天，我对自己完成的看似不可能的事情而感到惊奇。作为程序员，我得一行一行地写程序代码，我会把多行代码合并为叫做子程序的程序块，把程序又进一步合并成为程序模块，再将模块组合起来形成应用软件。当时，我着手编写的电子表格软件(spreadsheet program)是由大量的子程序和模块组成的，一般人很难读明白其中所有的代码。程序实在是太复杂了，而每一行代码的作用却微乎其微。在显示器上显示一个像素需要多行代码，在整个屏幕上画上电子表格需要计算机执行散布于几百个子程序中的上百万条指令，子程序又反复、递归地调用其他的子程序。程序确实太复杂了，程序一旦运行起来我们几乎就不可能知道将会发生什么。程序运行起来的片刻间，它就画好了需要画的图形，这让我觉得真是难以置信。计算机画好了由数字、标签、文字和图表组成的表格。这个程序已经具备了电子表格的功能。我知道计算机内部到底是怎么回事，中央处理器每次只执行一个简单的指令，很难想像计算机能在迷宫般的程序模块和子程序中飞快地找到并执行所有的指令。如果不是我来写这个程序的话，我肯定会说它根本就无法运行。我想，当时如果某个人有了在有着图形用户界面的计算机上运行电子表格应用软件的主意，当他把程序写在纸上给我看的话，我肯定会拒绝这一想法，并说这肯定实现不了。我会说完成一件事需要很长的时间。实际上，这样的看法太过于保守了，因为程序确实能运行。就在那时，我才意识到对于微处理器运算速度和分级设计效能，我的直觉不太正确。

对于新皮层的问题来说，这是一个值得借鉴的教训。大脑

的新皮层并不是由超高速的组件组成，其运作的规则也并不复杂，然而，它却确实有着一个分级结构，里面包含着上亿个神经元和上万亿的突触。如果我们觉得这么一个逻辑原理简单而数值上庞大的记忆系统能产生我们的意识、我们的语言、我们的文化、我们的艺术、还有这本书和科学技术是难以想像的话，我认为这其中的原因是，对于皮层容量以及它的分级结构的效能，我们的直觉不太正确。而新皮层就是这样工作的。它没有什么魔力。我们能理解它，正如我们能研制出计算机来一样，我们最终也能研制出与新皮层有着同样工作原理的智能机器来。



第七章

意识和创造力

Consciousness
and Creativity





由于预测与大量的人类活动有联系，所以每当我就我的大脑理论做报告时，听众总能很快就领会到它的重要性。他们会提出许多与此有关的问题：创造力来自哪里？意识是什么？想像力又是什么？我们是怎样区分事实与假象的？虽然这些问题并不是我研究大脑最主要的动机，但几乎所有的人都对此感兴趣。我并不想把自己装扮成这方面的专家，但有关智力的记忆—预测理论框架却能为此提供答案，并有助于对这些问题进行深入了解。在这一章里，我将讨论一些人们最常提及的问题。

动物有智力吗？

老鼠有智力吗？猫有智力吗？在进化过程中智力又是何时出现的呢？我喜欢上述问题，因为我发现它的答案令人惊奇。

到目前为止我对新大脑皮层所做的阐述以及它的工作原理都是基于一个非常基本的前提——世界是有组织、有结构的，因此是可以预测的。世间万物都有固定模式：脸上长有眼睛；眼睛里有瞳孔；火是烫的；重力使物体下落；门可以开关；等等。这个世界不是随意构成的，也不具有同质性。如果世界不存在组织结构，那么记忆、预测以及行为就失去了存在的意义。所有的行为，不论是人的，还是蜗牛的，是单细胞生物的，还是一棵树的，都是一种利用世界的结构来进行繁衍的方式。

想像一下生活在池塘里的单细胞生物。它们长着一根鞭毛，可以在水中游动，表面是一些能够发现营养物质的分子。因为并非池塘里所有地方的营养物质都是平均分配的，所以从细胞的一侧到另一侧营养物质的量就会有一个逐渐的改变。当这个单细胞生物在池塘中游动的时候，它就可以发现这种变

化，从而形成了此种单细胞生物关于世界结构的简单模型：这个生物利用它对化学物质的识别力来寻找营养富集的水域。我们可以说这个单细胞生物在做一种预测——沿着哪条路线游动才能找到更多的营养物质。这种预测会有记忆参与其中吗？答案是肯定的，记忆就存在于这个生物的 DNA 中。单细胞生物一生之中是不会学习如何发现并利用这种营养物质的渐进变化的，相反地，学习是发生在进化过程中，并存储在这种单细胞生物的 DNA 里。如果世界的结构突然改变，那么这种单细胞生物是不会学着去适应这种改变的。它不能改进自己的 DNA 以及由此而引起的行为。对这种生物来说，学习只能发生在一代又一代的进化过程当中。

那么，这种单细胞生物有智力吗？从我们对人类智力的日常理解来看，答案是否定的。但是，在由利用记忆和预测来更成功繁衍的生物所组成的连续体当中，这种生物处于最远端，以这种学术性更强的观点来看，答案又是肯定的。问题不在于说一些生物有智力，而另一些生物没有智力，所有的生物都在运用记忆和预测，只是它们运用记忆和预测的手段和复杂程度不同而已。

植物也通过使用记忆和预测去利用世界的结构。当树根植于土壤中或枝叶向天空伸展时，它就是在做一种预测——它基于自己祖先的经验预测在哪儿可以找到水和矿物质。当然，它这并不是思考，其行为也是无意识的，但它们利用世界结构的方法与单细胞生物是一样的。不同的植物种类都有一套彼此不同的行为方式，来利用世界结构中稍许不同的部分。

植物最后形成了交流系统，这种交流系统很大程度上是基

于化学信号的缓慢释放。如果昆虫损害了树的一部分，这部分就会通过它的维管系统向树的其他部分传递化学物质，从而引发保护机制，比如像产生毒素。通过这个交流系统，树显示出了更为复杂的行为。神经元可能是进化而来的一种交流信息的方式，不过这种方式比植物的维管系统更快捷。你可以把神经元想像成有着自己维管结构的细胞。在进化中的某一时刻，神经元开始使用生物化学电，而不是在这些维管中缓慢地传输化学物质，因为这种信息传递方式的速度快得多。起初，快速的神经轴突的传递和简单的神经系统并不涉及过多(如果有的话)的学习，关键在于这样传递信息更快。

然而，在进化过程中，出现了一些非常有趣的情况——神经元之间的联结可以改变。神经元可以根据最近发生的状况而决定是否发出信号。现在，在生物的生命周期内，行为可以改变，神经系统具有了可塑性。因为记忆可以很快地形成，所以动物在其一生中能够习得世界的结构。如果世界突然改变，比如说出现了一种新的掠食动物，其他的动物就不必坚持其基因所决定的行为，因为这些行为不再适合。具有可塑性的神经系统具有很强的进化优势，它可以引发大量新物种的出现，包括各种鱼类、蜗牛和人类等。

正如我们在第三章中所看到的，所有的哺乳动物都具有古脑，其顶部是新大脑皮层。新皮层是新近进化出来的神经组织。但正是因为有了层级结构、恒定表征以及通过类推而做出预测，新皮层能使哺乳动物比那些不具有新皮层的低等动物更会利用世界的结构。我们有了新皮层的祖先能够想出结网捕鱼，而鱼却不知道网意味着死亡，也不知道制造工具去割破渔

网。所有的哺乳动物，从老鼠到猫，再到人类，都具有新大脑皮层，它们都具有智力，只是程度高低不同而已。

人的智能有何不同？

关于这个问题，记忆 - 预测理论框架提供了两种回答。第一个回答是直截了当的：我们的新皮层比猴子或狗的大许多。我们的大脑皮层扩展到了一块餐巾的大小，因此我们能够了解更为复杂的世界结构，也能做出更复杂的预测。比起其他哺乳动物来，我们能理解更深刻的类比，更能了解结构之上的结构。我们在找对象时，不会只看他诸如身体健康与否之类的简单特征，而会与他的朋友、双亲面谈，观察他开车的习惯以及说话的方式，判断他是否诚实。我们就是这样通过第二位甚至第三位的属性来预测可能成为自己配偶的人今后的表现将会如何。股票交易者寻求各种交易模式中的结构特点，数学家寻找数字和公式中的结构，而天文学家则探索星体运动的结构。人类变大了的新大脑皮层使我们认识到自己的家是城镇的一部分，城镇又是地区的一部分，地区又是地球的一部分，而地球则是浩渺宇宙的一部分——结构之中又有结构。其他哺乳动物都没能思考到这样的深度。我深信我家的猫对我家房子以外的世界是没有任何概念的。

人类与其他哺乳动物相比第二个不同点就是人类具有语言能力。许多大本大本的著作都在讨论被认为是人类所独有的语言以及它是如何发展的。然而，语言不需要任何特别的辅料或是任何专门的语言机制就能很好地纳入记忆 - 预测框架。口头或书面的词汇，就如旋律、汽车和房子一样，都是世界上的模

式。语言的句法和语义与其他日常事物的层次结构并无差异。同样地，我们将火车的声音与火车的视觉记忆图像关联起来，将话语与客观实在及语义对等体的记忆关联起来。通过语言一个人可以唤起另一个人的记忆，可以在其头脑中建立心理对象之间新的关系。语言就是类比，通过它我们能够让其他人体验并了解他们也许从来都未曾看到过的事物。语言的发展需要一个能处理句法和语义嵌套结构的大的新大脑皮层，同时还需要一个更加充分发展的皮层运动区和肌肉系统，使我们能发出复杂的、高度清晰、连贯的语音或做出手势。我们可以利用语言将终生所学得的模式传给后代或部族。语言，不论是写出来、说出来，还是体现于文化传统之中，已成为世代传承我们关于世界知识的手段。今天，印刷品和电子通信使我们能与世界上各个角落的千百万人分享知识和信息。没有语言的动物无法传递给后代太多的信息。一只老鼠一生中能学到许多模式，但它却不会将细节的新信息传给后代——“嘿，小子，这就是家父教我怎样避免触电的高招。”

因此，智能的发展经历了三个时期，每个时期都利用了记忆和预测。

在第一个时期，生物将 DNA 作为记忆的载体。个体在其生命周期中不能学习和适应。它们只能将基于 DNA 的关于世界的记忆通过基因传给后代。

第二个时期始于自然界出现了能快速形成记忆的可改变的神经系统。此时，个体便能在它的生命周期内认识世界结构的重要方面并相应地调适自己的行为。但是，个体除了直接观察以外仍然不能将知识传给后代。在这个时期里，新皮层开始出现并扩大，但并不是这个时期主要的特点。

第三个，也是最后一个时期，是人类所特有的。这个时期是随着语言的产生和新皮层的扩展而开始的。人在一生中学会了世界结构的许多知识，并能运用语言有效地将这些知识传递给许多人。此刻，你我正在参与这样的过程。我花了生命中很长一段时间来探索大脑的结构以及这样的结构是如何导致思维和智能的产生的，通过本书，我将自己的所学传递给你。当然，如果我不能获取数以百计的科学家的研究成果，我是无法取得现在的成果的。而那些科学家们又是从他人的研究那里获得营养，一代又一代地将成果长期积累形成的，我不过是将他人的观点和观察进行消化吸收并融入自己的见解。

我们已成为这个星球上最能适应环境的动物，也是唯一能够在群体中广泛传播关于世界知识的物种。人口数量之所以激增，是因为我们能认识和利用世界众多的结构并将其传播给他人。在任何地方，无论是雨林、沙漠，还是寒冷的苔原，我们都能兴旺繁荣。扩展了的新皮层与语言的结合促进了人类持续的成功。

何谓创造力？

经常会有人问我有关创造力的问题。我想，正是因为许多人将创造力看成是一种机器所不具备的能力，因此这个问题就构成了对创建智能机器整个构想的一种挑战。那么，何谓创造力呢？这个问题的答案本书中已经涉及多次。创造力并不仅仅发生在大脑皮层的某个特定区域，它也不像情感和平衡感那样，根植于大脑皮层外部某个特定的结构或回路上。实际上，创造力是大脑皮层各区域所固有的一种属性，是预测的必

要组成部分。

如何理解这种说法?创造力难道不是一种需要很高智能和天赋的不同寻常的能力吗?其实不然。创造力可以被简单地看成是通过类推而进行预测的一种活动而已。它发生在大脑皮层的各个角落,只要人醒着,它就时时在发生。各种创造力构成了一个连续体,小到发生在皮层感觉区平常的感知行为(比如听到一首用新的调子演唱的歌曲),大到发生在大脑皮层最高领域的高难度、罕见的天才行为(比如以全新的方式创作交响乐)。从根本上说,它们极其相似,只不过这些平常行为太过于普通了,以至于我们没有注意到它们而已。

到此为止,对于我们是如何产生恒定记忆、如何运用恒定记忆来做出预测,以及如何预测出一些与我们过去所经历过的多少不完全一样的将来事件等问题,你应该有了一个基本的了解。可以说,我们的恒定记忆也是有关事件序列的记忆。通过将此刻发生的事件细节与下一刻将要发生事件的恒定记忆结合起来,我们便能做出预测。预测就是将恒定记忆序列应用于新的情况,因此脑皮层的所有预测都是通过类推而做出的。我们通过类比过去而预测未来。

假设你准备在一家不熟悉的餐馆用餐,你想先洗一下手。即使你从未来过这里,你的大脑也会预测到这家餐馆的某个地方会有一个装有洗手用的盥洗盆的洗手间。你的大脑是如何知道的呢?你曾去过的其他餐馆有这样的洗手间,通过类推,这家餐馆很可能也有一个类似的洗手间。而且,你知道要上哪儿去找它,它会是什么样子的。你预测到会有一个门或标志,上面有某种表示男女的符号。你预测到它将会在餐馆的后面,要么

是在柜台一侧，要么在厅堂一角，一般情况下它不会设在就餐区视线所及的区域内。总之，虽然你从未来过此餐馆，但通过类比其他餐饮店，你就能找到你所需要的。你不会毫无头绪地四处乱找，你会去寻找那些能尽快找到洗手间的预期中的模式。这种行为就是创造性行为，即通过类比过去而预测未来。在一般情况下我们会觉得这再正常不过，而这类行为其实就是创造性行为。

最近我买了一架颤音琴(vibraphone)。我从未弹过颤音琴，但我们原先有一架钢琴。颤音琴买回来的当天，我就从钢琴上取了页乐谱，放在颤音琴的谱架上，然后开始弹奏起简单的曲子。这样的能力当然不算什么，但从本质上讲，这就是创造性行为。颤音琴与钢琴的外观、构造和演奏方式有很大的不同：这种乐器的琴键是金黄色的金属键，而钢琴的琴键却分黑、白两种；颤音琴的金黄色琴键比较大，每个键的大小都一样，钢琴的琴键却比较小，还分两种大小；颤音琴的琴键排成两排，而钢琴的琴键却是黑白相间，排成一行；是钢琴用手指弹奏，而弹奏颤音琴却是挥动木锤；颤音琴要站立弹奏，而弹钢琴却是坐着；况且弹奏颤音琴所涉及的肌肉和动作与弹钢琴的截然不同。

那我为何能在一个架根本不熟悉的乐器上进行弹奏呢？答案就是，我的大脑皮层看到了钢琴琴键与颤音琴琴键的相似之处。正是利用这种相似性，我才得以弹奏曲子，这就好比换一个调子唱同一首歌一样并没有本质的不同。我们都知道如何利用过去的知识对上述情况进行类推。我知道这两种乐器之间的相似性对你来讲是显而易见的，但那正是我们的大脑能自动地

进行类推的结果。如果试图编个程序让计算机来寻找诸如钢琴和颤音琴等物体之间的相似性的话，你会发现那将是何等的困难！创造力——利用类推而进行预测——真是无处不在，而我们一般都不会注意到它。

然而，我们认为只有当记忆-预测系统在做高层次的抽象活动时，也就是当它利用非同寻常的类推而做出非同寻常的预测时，此刻的行为才是创造性的。例如，大多数人都会一致认为数学家证明了一个艰难猜想是具有创造力的。但是，让我们仔细看看她的心理活动都涉及了些什么。数学家凝视着等式，自言自语道：“这个问题该如何解决？”如果答案很难看出，她会重列等式。通过另一种方式来表述问题，她便可以从一个全新的角度来审视同一个问题。她继续凝视着。突然她发现等式的某一部分很眼熟。她想：“哦！这个我知道。这个等式的一个部分与多年前曾列过的一个等式有相似之处。”这就是在利用类推而做出预测。“或许我可以利用曾成功解出过去那个等式的方法来解出这个等式。”她能通过类推以前成功解决问题的方法来解决现在的问题。这就是创造性行为。

家父曾患过一种奇怪的血液病，连医生开始都束手无策。那他们最终是如何给出治疗方案的呢？他们所做的一件事情就是查看那些逐月从分析家父血液而获得的数据，目的是看从中能否识别出有效的模式来。（家父把那些数据绘成美观的图表以便让医生能看得更清楚。）即使病症与那些已知的疾病症状并不完全吻合，医生们也找到了某些相似之处，他们最后的治疗方案是在综合了治疗其他血液病的方案之上而确定的。治疗方案是根据类比他们曾治疗过的疾病而做出的猜测。要能识别出

这些模式需要对疑难杂症见多识广才行。

莎士比亚的比喻是创造力的典范。诸如“爱情是叹息吹起的一阵烟(Love is a smoke made with the fume of sighs.)”“用哲学的甘乳安慰你的逆运(Adversity's sweet milk, philosophy.)”“人们的笑脸里都暗藏着利刃(There's daggers in men's smiles.)”这样的隐喻读起来让人容易明白其意味，但却是常人难以创造出来的。这也是为什么莎士比亚被认为是文学天才的原因之一。要创造出这样的隐喻，他得能看出事物间一系列巧妙的相似性才行。当他在写“人们的笑脸里都暗藏着利刃”时，莎翁并不是在说利刃或笑脸。利刃与坏心眼相类比，而人们的笑脸则与骗术相类比。两个如此巧妙的类比仅用了5个字！至少这是我的理解。诗人的天赋在于：将看似毫不相干的词语或思想联系起来，并用这种方式照亮世界。他们创造出让人意想不到的类比，能使人们借此认识到更高层次的结构。

事实上，富有创造性的艺术作品之所以得到大家的喜爱，是因为它们出乎我们的预想。当看到一部影片打破你熟悉的人物模式、故事情节以及摄制技术(包括特技在内)时，你会因为它不落俗套而欣赏它。油画、音乐、诗歌、小说等所有这些创造性艺术形式——都在努力打破传统的框架，并尽可能要出乎受众的预料。在此，矛盾的力量使得艺术伟大：一方面我们要艺术耳熟能详，一方面又要求它与众不同、出人意料。太过熟悉是重复和俗气；太过于独特则显得不协调和难以理解。最好的作品就是打破我们期望的模式并且同时教给我们新的模式。想想伟大的古典音乐吧！最好的音乐在最基本的层面上——好的节奏、简单的旋律和乐句——具有吸引力，所以，任何人都能

理解并欣赏它。然而，其中仍有少许不同和意外。但你听得越多，就越能感觉到在出人意料的部分有着某种模式，比如说重复出现的非同寻常的和声或调式变化。伟大的文学作品或电影也不例外。你越是读得多或看得多，就越能体会到结构中细节的创造性和复杂性。

你也许有过这样的经历：当你看到某个事物时，突然间你想不起来它是什么了。“嗯，这个模式我似乎在什么地方见过。”此刻，你并不是在试图解答难题，只不过是新的场合在激活你大脑中的恒定表征。你看到的是一般情况下毫不相干的两种事物间的相似之处。我也许可以这样认为：推进一个科学思想可与出售一个商业点子相类比；引发社会变革可与抚育孩童相比拟。假如我是诗人——啊，难道不是吗？——我创造了新的隐喻。如果我是科学家或者工程师，则为一个长期困扰我们的难题找到了解决之道。创造力就是将过去曾经体验过的模式与未来生活将要认识的模式彼此糅合、匹配。大脑这样说：“这个不就是有点像那个嘛。”此种神经机制在大脑皮层中无处不在。

某些人比其他人更富有创造力吗？

我们经常会听到这样的问题：“如果所有人的大脑都天生具有创造性，为什么我们的创造力会存在差别呢？”记忆-预测理论框架提供了两种参考答案，一种与环境因素有关，另一种与自然属性有关。

从环境方面看，每个人都有自己不同的生活体验，因此每个人在自己的脑皮层里形成了对世界的不同模型和记忆，而且做出的类推和预测也不尽相同。如果我接受过音乐训练，就能

够用新的调子唱出一首歌来，也可以在新的乐器上弹出简单的旋律。如果从未接触过音乐，我显然不可能做出这些预测性的新行为。如果学过物理，我就能够从物理规律类推并解释平常事物的运动特性。如果和狗一起长大，我就能对狗的习性了如指掌，并能很好地预测它们的行为。一些人在社交场合、语言、数学或外交手段方面更具创造性，这些都归因于他们成长的环境。我们的预测，进而是我们的天分，都是在体验的基础上得以发展的。

在第六章中，我描述了记忆是如何在脑皮层体系中形成的。据此可知，某种模式你接触得越多，在脑皮层的低层级上这些模式的记忆就被改造得越多。它会使你在脑皮层的高层级中理解更为抽象的事物之间的关系，这就是所谓“专业”的实质。专家就是这样的人，他们通过练习和反复的体验，便能识别那些非专业人士无法识别的微妙模式。例如，他们能识别 50 年代后期小车上的鳍状物，海鸥喙上斑点的大小。专家能识别模式之上的模式。最终，由于受到大脑皮层大小的限制，我们的识记能力有一个生理限制。但作为人类，与其他物种相比，我们的脑皮层是阔大的，因此，我们所学的知识是丰富的，其丰富程度取决于我们一生中所接触的事物。

在自然属性方面，大脑表现出了自然的差异性。当然，有些差异是由基因决定的，比如区域的大小（V1 区的面积个体差异就能达到 3 倍），以及半球侧化（连接左右半脑的神经，女人比男人的要粗厚）。人脑中细胞的多少与连接类型似乎也有所不同。阿尔伯特·爱因斯坦的创造性天才不太可能是完全因为他年轻时曾工作过的专利局所具有的激励性环境造就的。最近对

他大脑(曾一度被认为遗失了,后来,也就是在几年前被发现保存在一个罐子里)的分析表明,他的大脑经测量显示出了不同寻常的特点:他的大脑与普通人的相比,有更多神经胶质的支撑细胞;在顶叶中——被认为对数学能力和空间推理能力有着重要作用的区域,他的脑沟(或脑槽)有着与众不同的模式;比起大多数人的脑子来,他的大脑也要宽上 50%。我们也许永远也无从知道爱因斯坦为什么会如此聪明,如此具有创造天赋,但有一点是可以确定的,那就是他的天才有一部分是来源于遗传因素。

无论异常聪明的大脑与普通的大脑之间有怎样的差异,我们都是富有创造性的,通过练习和学习,我们都能提高自己的技能和才智。

能把自己训练得更具创造力吗?

是的,的确可以。我已经发现,在解决问题时有一些方法能培养我们找到有用的相似点。首先,你要设想自己正在面对的问题是有答案的。人们容易轻言放弃,但你要相信问题的答案正等着你去发现,你要为此而坚持进行长时间的思考。

其次,你要让自己的思维任意驰骋。给你的大脑充分的时间和空间来发现问题的答案。找到一个问题的解答实际上就是要找到存储在大脑皮层中与当前要解决的问题相似的模式。如果你对某个问题感到困惑,记忆-预测模型建议你从不同的角度去观察它,这样会提高用以往经验找到相似点的可能性。如果你只是静静地坐在那儿,一遍又一遍地看着问题,那是不会有任何进展的。试着把问题分解一下,将问题的各个部分用各

种不同的方式进行重新组合。当我玩拼字游戏时，我会不断地变换它们的顺序。我并不是希望碰巧用一些字母凑出一个单词，而是希望不同的字母组合可以让我想起可能是问题答案的某个单词或这个单词的一部分。如果你正看一幅看不明白的图画，试着将它上下颠倒，或者变换颜色，或者变换角度来看。比如说，当我琢磨 V1 区中不同的模式是如何产生 IT 区中恒定表征的时候，就被卡住了。因此，我将问题的顺序颠倒一下，也就是考虑 IT 区的稳定模式是如何产生 V1 区的不同预测的。将问题颠倒过来便即刻奏效，这使我相信 V1 区不应被看作是单一的脑皮层区域。

如果在某个问题上你被卡住了，那么暂时将它放到一边，先做点别的事情，然后再重新开始，将问题改变一下提法。如果你这样反复多遍，可能要花上几天或几个星期的时间，但问题迟早会有些眉目的，你会在过去的经验中找到相似的情形。要想成功，就要反复琢磨问题，同时还要做点其他事情，这样，大脑皮层就会有可能会找到一些相似的记忆。

下面的例子可以说明对问题进行调整是如何有效地帮助我们找到全新解答的。1994 年，我和同事正试图找到在掌上电脑上输入文本的方法。大家都将注意力放在手写识别软件上，他们说：“你看，你在纸上写东西，你同样可以在电脑显示屏上书写。”可惜的是，这样做太难了。这对人脑来说虽算不上什么，但电脑却不擅长此道，这是因为人脑利用记忆和上下文便能预测写的是什么。割裂开来无法识别的单词或字母在上下文环境里就很容易被识别。电脑的模式匹配不足以解决这个问题。我设计了几种利用传统手写识别技术的软件，却都不令人

满意。

我花了好几年的工夫想将识别软件做得更好，但却没有什么进展。一天，我又回到问题的起点，决定换个角度重新审视这个问题。我一边思考与此类似的一些问题，一边在想：在台式计算机上我们是怎样输入文本的呢？我们是用键盘来键入的。那我们又是如何掌握在键盘上打字的呢？实际上，那并不容易。在键盘上打字是最近的发明，并且需要花较长时间来掌握。在类似于打字机的键盘上打字是比较困难的，也不直观，与书写很不一样，然而，数百万的人都在学如何打字。为什么会这样呢？因为学会打字是很有用的。我的思路继续以类推的方式展开：也许我能发明一种文本输入系统，它不必很直观，也需要学习，但人们会接受并使用它，因为它很有用。

从字面上讲，那就是我思考的过程。我用在键盘上打字作类比，想找到用触笔在显示屏上键入文本的方法。我意识到人们是愿意去学习一个困难的任务(打字)的，因为这是一种将文本输入机器的可靠且快速的方法。因此如果我们能够发明一种用触笔既快捷又可靠地键入文本的新方法，即使需要学习，人们也是会采用它的。为此我设计了一种字母，能可靠地把你手写的内容转换为电脑文本；我们把输入方式叫做 Graffiti(即“涂鸦手写识别系统”)。当使用传统的手写识别系统时，即使计算机认错了你写的字，你也无从知道那是怎么回事。但 Graffiti 系统总是能识别出正确的字母，除非是你自己写错了。人脑讨厌不可预测的东西，这也是为什么人们不喜欢传统的手写识别系统的原因。

许多人认为 Graffiti 是一个非常愚蠢的想法，因为它违反了



他们所认为的计算机应有的工作方式。当时的口号就是计算机要适应用户，而不是要求用户来适应计算机。但我有理由相信人们会接受这种类似用键盘打字的新的文本键入法。Graffiti 已被证明是一个很好的解决方案，而且得到了广泛应用。可惜直到今天，我还能听到有人宣称计算机要适应用户，其实这并不一定有道理。我们的大脑更喜欢一贯稳定且有预测性的系统，而且我们还总是喜欢掌握新的技能。

创造力会使我误入歧途吗？我会被自己所愚弄吗？

错误的类推总是危险的。科学史上不乏这样的例子，一些绝妙的类推，最后被证明是谬误。例如，著名天文学家约翰尼斯·开普勒(Johannes Kepler)认为 6 大已知行星的轨道是由正立方体确定的。正立方体是完全由正多边形组成的唯一立体形状，总共有 5 种正立方体：正四面体(由 4 个等边三角形组成)、正六面体(由 6 个正方形组成，也叫正方体)、正八面体(由 8 个等边三角形组成)、正十二面体(由 12 个正五边形组成)和正二十面体(由 20 个等边三角形组成)。它们是由热衷于数学和宇宙之间关系的古希腊人发现的。

如同所有文艺复兴时期的学者一样，开普勒深受古希腊思想的影响，对他来说，宇宙中有 5 种正立方体和 6 大行星不可能是巧合。正如他在《宇宙的神秘》(1596 年)中所说的：“运动着的世界是由正立方体所表示的。这些立方体一共有 5 个，当把它们看作界线时，它们决定了 6 个不同的物体——绕太阳运行的 6 个行星。这也是为什么只有 6 大行星的原因。”他所做的类推很漂亮，但却是完全错误的。

开普勒继续根据以太阳为中心、相互嵌套的立方体来解释行星的轨道。他以水星轨道所确定的球形为基线，在它的外面画了一个八面体，这个八面体的顶点确定了一个更大的球形，即金星的轨道。在金星轨道以外他又画了个二十面体，其顶点构成了地球的轨道。这一系列推断继续往下：地球轨道外的十二面体构成火星的轨道，火星轨道外的四面体构成木星的轨道，木星轨道外的正方体构成土星的轨道。这样的类推真是简单又漂亮。鉴于当时的天文学数据精确性有限，他才会确信这个宇宙蓝图是成功的！[几年后，当他获得已故的同事第谷·布拉赫(Tycho Brahe)更为精确的数据时，开普勒才认识到自己错了。这些精确的数据证明行星轨道是椭圆的，而不是圆形的。]开普勒的趣闻给科学家以及所有的思想家敲响了警钟。大脑是建立模型、做出创造性预测的器官，但它建立的模型和做出的预测可能是正确的，同样也可能是个美丽的“谎言”。我们的大脑总是在考察模式，做出类比。如果事物间正确的联系没能被揭示，我们的大脑会很满足地接受错误。伪科学、偏见、宗教还有偏执狭隘经常根植于错误的类推之中。

意识是什么？

这是让神经科学家头痛的问题之一，在我看来，大可不必如此。有些科学家，如克里斯托弗·科赫，愿意研究有关意识的问题，但大多数的科学家却认为这是一个接近伪科学的哲学命题。我认为这个问题仅仅就凭很多人对此好奇这一点，就值得深入探讨。对这个问题我提供不了一个完美无缺的答案，但我以为，记忆-预测模型却能部分地回答这个问题。

不久前，我参加了一个在美国迷人的长岛海湾召开的科学会议。还不到傍晚，晚饭还没开始，晚间还有会议，我们十几个人端着酒杯，沿着码头走到水边，开始闲聊。一会儿，话题就转到了意识这个问题上。正如我所说的，神经科学家一般不会谈论这个问题。但因为我们身处美景，又小酌了几杯，很自然地就向这个话题进军了。

一位英国科学家大谈她关于意识的观点，她说：“我们永远也不会理解意识。”我反驳道：“意识其实不是一个难题。我认为意识不过是脑皮层的感受。”开始时所有人还都沉默不语，紧接着讨论展开了。有几个科学家试图纠正我的观点，要指出我的明显错误：“你必须承认世界充满了活力，前景光明。你如何能否认那能感知世界的意识？你得承认你感到了异样的东西。”为了支持自己的观点，我说道：“我不明白你在说什么。鉴于你谈论意识的方式，我不得不得出结论——我与你不同。我感觉不到你所感觉到的东西，所以大概我不是一个有意识的动物，我一定是一具僵尸。”当哲学家谈论意识时，他们通常都会提到僵尸。僵尸与人都有肉体，但僵尸却没有意识，它们是没有家园的行尸走肉。

那位英国科学家望着我，“你当然有意识。”

“不，我不这样认为。在你看来我可能是有意识的，但我并不是一个有意识的人。对此你不必担心，我很正常。”

“难道你没感受到这美景吗？”她指着那波光粼粼的海水说。此时，夕阳西下，天空一片橙红。

“是的，我看到了这一切。那又怎样？”

“那你如何解释你现在的主观感受呢？”

我答道：“是的，我知道我站在这里。我头脑中有类似今晚情形的记忆，但我并不觉得有什么特别。因此，对此如果你觉得有什么特别的话，那我大概就是没有意识了。”我尽力使她明白她的想法非常奇妙，并努力让她给意识下个定义，但她却不能解释意识。

我们一直持续着这场论战——一个说我有意识，一个说我没意识——直到得动身去就餐了。我想我并没有改变任何人对意识以及它存在的方式和意义的看法，但我却努力让他们认识到这样的现实，即多数人把意识当作是某种神秘的浆汁，它被加到了大脑实体之上。你拥有一个由细胞组成的大脑，然后你把意识，这种神奇的浆汁注入其中，那就是人现在所处的情形。根据这样的观点，意识是与大脑分离的某种神秘实体。这也就是为什么僵尸有大脑但缺乏意识的原因。僵尸有所有应有的物质原料、神经元、轴突，但没有那特殊的浆汁，它们能做人能做的一切。从外表上看，你无法区别一具僵尸和一个人。

认为意识是特殊物质的观点源于早期关于“元气(elan vital)”的信仰——一种曾被认为使生物具有活力的神秘力量。人们曾以为我们需要这种生命力才能解释岩石与植物、金属与少女的区别所在。现在，已几乎没有人相信此种观点，因为我们对生命体和无生命体之间的差异了解甚多，也就不再认为生命力是一种特殊的浆汁了。现在我们对 DNA、蛋白质折叠、基因转录以及新陈代谢有了深入的了解，虽然我们对生物体的整个机制还不能完全掌握，但现有的生物学知识已将神秘之力排除在外了；同样地，人们不再认为肌肉运动需要魔力或精神，我们已知折叠的蛋白质能将长分子交叠在一起。这些你都可以从

书本上了解到。

然而，许多人仍坚持认为意识与众不同，它不能从还原理论的生物学观点得到阐释。再次申明，我不是研究意识的专家，也没有读过所有哲学家的观点，但我却能知道在这场论战中他们有所混淆。我认为意识就是有了新大脑皮层后的感觉。其实，除此之外我们还可以了解更多。例如，我们可以将意识分为两大类：第一类类似于自我意识，也就是平常我们所说的有意识。这一点比较容易理解。第二类就是可感受性(qualia)——认为与感觉关联的情绪独立于感觉输入。而第二类才是更难理解的部分。

当大多数人提到“意识”这个词时，他们指的是第一类意识。“你有没有意识到从我旁边经过时，你连招呼都没打呀？”

“昨晚从床上滚下来时，你有没有意识到？”“睡觉时你是没有意识的。”有人说这种意识就是知觉，两者虽然很接近，但我认为知觉并没有很好地解释意识。我认为这时的意识类似于形成陈述性记忆。陈述性记忆是你能回忆起来并告知他人的记忆，并可以用语言来表达清楚。如果你问我上周末去哪里了，我可以告诉你。那就是陈述性记忆。如果你问我如何平衡自行车，我会告诉你手握自行车扶手，脚蹬踏板，但我却不能确切地解释如何做。如何平衡自行车主要与古脑里的神经活动有关，所以它不是陈述性记忆。

一个小小的思维实验能告诉我们，日常理解的意识就是陈述性记忆。我们认为记忆是神经元及连接神经元的轴突之上的物理变化，还记得吗？因此，如果我有一种方法能将这种物理变化反转，那么你的记忆就会被抹掉。现在想像一下，我能拨动

一个开关，将你的大脑恢复到过去的某个时间点上的状态，可以是1个小时前，24个小时前，或者任何其他时间。我只是拨动回溯机器上的开关，你大脑中的神经元和轴突就回到了先前的某个状态。这样我就抹掉了从那一刻起你所有的记忆。

假设你度过今天，明天醒来。但就在你醒来之际，我拨动了你头脑中的开关，抹掉了你过去24小时的记忆。从你大脑的角度来看，你应该没有了前一天的记忆，因为昨天根本就不存在。我告诉你今天是星期三，那么你会反驳道：“不对，今天是星期二。我能确信，日历肯定被多翻了一页，今天绝对是星期二。你为什么要捉弄我呢？”但所有在星期二见过你的人都会证明一整天你都是有意识的。他们看到过你，与你共进过午餐，也与你交谈过。你难道不记得了吗？你肯定会说，不对，这绝没发生过。最后，他们又拿出你吃午餐时的录像，你才逐渐相信即使你一点也不记得所发生的一切，那天的事确实发生过。似乎你一整天就像具僵尸，没有意识。然而，当时你是有意识的。就是在你的陈述性记忆被抹掉时，你对自己有意识的看法随之消失了。

我们从这个思维实验可以了解到，陈述性记忆就是我们平常理解中的意识。如果在打完两小时的网球以后我问你对打网球是否有意识，你一定会说当然有。但如果我把你最近两小时的记忆抹去，你一定会说那段时间里自己没有意识，并且说那期间所有的行为都不是自己所为。在上述两种情况下，你都打过同一场网球比赛，唯一的不同就在于当我问你的时候，你对这场球赛是否存在记忆。因此，这里我们所说的意识不是绝对的，它会随记忆的被抹掉而改变。

关于意识，更难理解的就是可感受性的问题，这个问题有点像问禅说禅。比如，红色为何为红色，绿色为何为是绿色？红色在我看来和在你看来是否一样？为什么红色会给人带来某种感受？红色对我来说有一种无法摆脱的感觉，那么它对你又会怎样呢？

我觉得这样的描述难以与神经生物学挂起钩来，因此，我要将问题换个提法。还是同一个问题（这个问题我也无法回答），那就是：为什么不同的感觉看上去确有本质上的不同？为什么视觉不同于听觉，听觉又不同于触觉？如果说脑皮层的各个部分构造都是一样的，如果说它的工作过程也是一样的，如果说大脑不过是在处理模式，如果说进入大脑的既不是声音也不是光，而全都是模式的话，那么为什么视觉会与听觉迥异？我觉得要描述视觉与听觉的不同是相当困难的，但它们很显然存在着差异性。我想对此你也有同感。然而，脑皮层中代表“声”和“光”的轴突实际上是完全一样的，沿着感觉神经元的轴突传递的并不是“光”和“声”。

得联觉病症的患者，他们的大脑混淆了各种感觉——某些声音有颜色，某些纹理也有颜色。这告诉我们感觉的本质并不是恒定不变的。通过某种物理变化，大脑可以使听觉输入具有视觉的某些特性。

那又如何解释可感受性呢？我想到了两种可能的解释，但都不完全令人满意。其中一个就是，虽然听觉、触觉和视觉在新大脑皮层工作的原理相似，但它们在新皮层之下被处理的方式不同：听觉依赖于一组专门针对听觉的新皮层下组织，它们在听觉模式到达脑皮层之前要先行处理；体觉模式也要先经过一组新皮

层下专司体觉的区域。也许可感受性与情绪一样并不仅仅是由新大脑皮层来调节的。如果说可感受性以某种方式与新大脑皮层下组织相连，这些组织有着特殊的联结方式，也许还与情绪中心相连，这样可能解释为什么我们会有不同的感觉，虽然这并不能帮助我们首先解释为什么有各种不同的可感受性感觉。

另一个可能的解释就是输入的结构——模式本身的不同，规定了你该如何去体验信息的性质。听神经上时空模式的本质与视神经上时空模式的本质是不同的。视神经有百万条纤维，能传递大量的空间信息。听神经只有3万条纤维，因而传递更多的是时间信息。这些差异也许与我们所说的可感受性有关。

可以肯定的是，不管意识如何界定，记忆与预测对于它的产生起着决定性的作用。

与意识有关的另外两个概念是心智和灵魂。

孩提时我曾想知道，如果“我”出生在另一个国度的另一个孩子的身体里，就好像“我”在一定程度上独立于我的身体之外，那将会是怎样的一种情况呢？这种心智独立于身体的感觉非常普遍，它是新大脑皮层工作的结果。脑皮层在其分级结构记忆中建立了一个有关世界的模型，思维就是这个模型自行运转产生的结果：记忆引起预测，预测又成为感觉输入，引起新的记忆，如此反复。我们大多数的冥思苦想并不是由现实世界驱动的，甚至不与现实世界直接相连，它们不过是我们头脑中世界模型的产物。我们闭上双眼寻求安静，是为了使我们的思考不被感觉输入打断。诚然，我们头脑中已建立的模型最初是通过接触现实世界，经由我们的感觉产生的，但当我们对世界进行规划和思考时，我们是通过脑皮层中的世界模型而不是世界本

身来完成的。

对于新脑皮层而言，我们的身体是外部世界的一部分。请记住，大脑处在一个安静的暗盒里，它只能通过感觉神经纤维上的模式来了解世界。我们把大脑看作是一个模式处理装置，从这一角度看，它对我们身体的了解与对世界其他部分的了解没有什么两样，因为身体与世界间没有特别的界限。但新脑皮层却不能建立关于自己的模型，因为大脑里没有感觉。因此，这就是为什么我们会觉得思想独立于身体，为什么我们会觉得有独立的心智和灵魂的原因了——脑皮层建立了有关你身体的模型，但它却不能建立有关它自身的模型。你的思想，位于大脑之中，从物理性质上讲，它与身体和世界是隔离开来的。心智独立于身体，但它并没有独立于大脑。

从创伤和疾病的比较，我们可以清楚地看到这种差异。如果一个人失掉了一只胳膊，他大脑中关于这只胳膊的模型依然完好无损，结果产生了所谓的“幻肢”，也就是他仍然觉得已失去的胳膊附在自己的身体之上。但反过来，如果胳膊仍在，而是脑皮层受了伤害，他就可能失去了大脑对这只胳膊的模型，这种情况下，他可能会罹患异肢症，从而产生一种不舒服、也许无法忍受的感觉，觉得这只胳膊不是自己的，而是在受他人的控制。如果我们的大脑完好无损而身体的其他部分生病了，我们就会觉得身体将死会导致健康的心智受累，虽然事实上是身体将死而受累的是健康的大脑。这样，我们很自然就会认为我们的身体死后心智仍继续存在，但事实上，大脑死亡后，心智也就随之消失了。如果我们的大脑先于身体死亡，这其中的道理就会变得更加明白。患有阿尔茨海默氏病或有严重

脑损伤的人，即使身体依然健康，也会失去心智。

什么是想像力？

从概念上讲，想像力是相当简单的。传入各皮层区的模式不是来自感官就是来自记忆层级中较低的区域。每一皮层区都产生预测，预测沿着层级向下传递。当你在进行想像时，你所做的仅仅是让预测调转回来，成为新的输入。你不需要实际去做什么事，预测的结果就可以生效。“如果这样的话，结果就会这样，接下来就会这样。”如此不断。每当准备商务会谈、或象棋对弈、或备战体育赛事等诸多其他事务时，我们都是这么做的。

下象棋时，你会想像把马移到某个位置上，并设想此时的棋局将会是什么样子。你根据心中所想的棋局的对阵，来预测对手将如何反应，棋局又会有什么变化，然后你又将如何应对，等等。你在想像中一步一步地下着棋，并思考着每一步的后果。最后，依据这一系列棋子的走法，你会判断出起初的一步是不是一着好棋。有些运动员，比如说滑雪运动员，如果他们在心中反复演习比赛的滑道的话，就能够提高成绩。闭上双眼，想像每一个弯道，每一个障碍，甚至站在领奖台上的情形，这对于他们增加成功的机会是会有帮助的。想像实际上就是策划的另一种说法，它是脑皮层预测能力所起的作用——允许我们在实际行动之前就能知晓后果如何。

想像需要一种神经机制来将预测转变为输入。在第六章中，我提出第6层中的细胞是负责精确的预测的，这一层的细胞将信息向下投射到层级中的较低区域，同时也把信息向上投

射到第4层的输入细胞，从而一个区域的输出就成为了自己的输入。长期从事脑皮层研究的斯蒂芬·格罗斯伯格把这种想像力的回路叫做“折叠反馈”。如果你闭上眼睛，想像一匹河马，脑皮层的视觉区就会活跃起来，就好像你真正看到了河马那样。也就是说，你看到了你所想像的事物。

什么是现实？

人们总是带着担心和惊讶问道：“你的意思是说我们的大脑建立了一个关于世界的模型，并且这个模型比现实世界更为重要吗？”

“啊，是的。在某种程度上说就是这样的。”我会这样回答。

“那世界难道不是存在于我们的头脑之外吗？”

的确是这样的。人是现实存在的，树也一样，我家的猫也是一样，你所处的社会地位也是一样。而你对世界的了解以及你对它的反应却是以你头脑内部模型的预测为基础的。在任何一刻，你只能直接感知这个世界的一小部分，这一小部分支配大脑来唤醒与此有关的记忆，但仅靠这一小部分却不足以建立你当前感知的完整内容。比如说，现在我在办公室里打字，听到有人敲前面的门。我早已知道我母亲要来看我，而我也认为她已在楼下，尽管我并没有实际看到她或听到她的声音，没有什么特定的感觉输入是与我母亲联系在一起的，是我关于世界的记忆模型通过类推过去的经验预测到她已经来了。你所感知的大部分并不是来自你的感官，它是由你内在的记忆模型所产生的。

因此，对于“什么是现实”这个问题的回答，很大程度上取决于

我们大脑皮层建立的模型能在多大程度上准确地反映世界的真实情况。

我们周围世界的许多方面都是稳定一致的，因此几乎每个人建立的内部模型都是一样的。你在婴儿时期就知道了光照在圆形的物体上会产生阴影，并且知道通过自然界的提示你就能判断出大多数物体的形状。你还知道如果从吃饭的餐台上将杯子扫下来的话，重力会使它掉落到地板上。你掌握了物体的纹理结构、几何知识、颜色以及日夜交替的规律。所有人学到的有关世界的基本物理属性都是一致的。

但我们头脑中世界模型有很大一部分是基于风俗习惯、文化传统以及父母的教育。我们大脑模型中的这一部分就不会那么一致了，也有可能是决然不同的。一个在充满仁爱和关怀的家庭，父母会很关心孩子的情感需求，当孩子长大以后可能会认为世界是安全而充满友善的；而那些受父亲或母亲、甚至双亲虐待的孩子，不管后来别人对他有多好，都会觉得未来生活危险而残酷，不会相信任何人。心理的许多方面都是早期生活经历、父母关爱和培育的结果，因为这正是大脑开始形成世界模型的重要时期。

你所处的文化氛围也完全决定了你的世界模型的样子。比如说，有研究表明，亚洲人和西方人对空间和物体的观察方式是不同的——亚洲人会更多地关注物体间的距离，而西方人更多的是关注物体本身。这种差异导致了不同的美学观和解决问题的不同方式。有研究指出，像阿富汗某些部落和南美的一些部族的文化，是建立在一种荣誉感之上的，结果，他们比较倾向于认为暴力是很自然的。早期养成的不同宗教信仰会导致完

全不同的道德观，诸如怎样对待男人和女人，甚至对生命本身的价值观都不一样。很明显，从绝对和普遍的意义上讲，这些关于世界的不同模型并不一定全都正确，虽说对个人而言它们可能是正确的。道德推理、是非观念都是后天学得的。

你所处的文化背景(包括家庭教育)教给你了一些典型，这些典型不幸成为了你生命中不可避免的一部分。在这整本书里，你可以将“典型”一词替换成恒定记忆(或恒定表征)，意思不变。通过类推而进行的预测，基本上就是基于恒定记忆所作的判断。消极的恒定记忆会带来可怕的社会后果。如果说我关于智能的理论是正确的话，我们就不可能消除人们用恒定记忆来思维的习惯，因为恒定记忆就是大脑皮层工作的本质。恒定记忆化是大脑的固有特点。

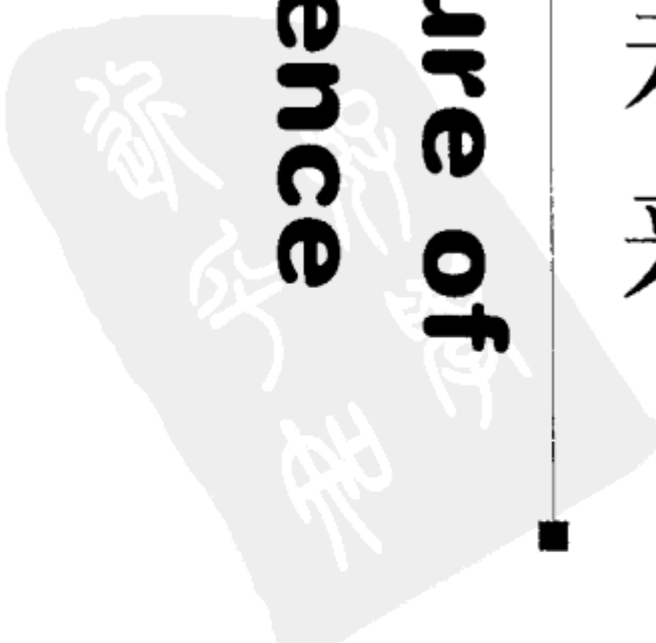
要消除恒定记忆引起的危害，其途径在于让孩子们能认识到错误的恒定记忆具有移情作用，是不可信的。在培养我们所认为的优秀的价值观的同时，我们还要提高他们批判思维的能力。怀疑论，作为科学方法的核心，是我们区分事实与虚构的唯一途径。

说到这里，我希望你已经相信，心智不过是大脑工作的标志，它不是操控大脑细胞或与大脑细胞共存的某种独立的东西。神经元就是细胞。没有什么神秘的力量会改变单个或多个神经细胞的正常工作方式。基于这一事实，我们可以将注意力转向如何将大脑细胞的记忆和预测能力——大脑皮层的通用算法——在硅晶片中实现了。

第八章

智能之未来

The Future of
Intelligence





我们要预言一项新技术的最终用途是困难的。正如大家在整本书中看到的，大脑是通过类比过去而做出预测的。因此，我们自然而然地认为，新技术仅仅是被用来完成那些旧技术所要完成的同样任务，只是更快、更高效、更经济。

这样的例子比比皆是。人们把火车叫做“铁驹(iron horse)”，把汽车叫做“无马马车(horseless carriage)”。在电报盛行的那几十年里，电话仅仅是扮演着电报的角色——只用于传递重要新闻和紧急事件。直到 20 世纪 20 年代，电话才得到普及。摄影问世之初被人们当作是一种新的肖像画技术。同样，人们视电影为舞台剧的变种，这就是为什么在 20 世纪的大部分时间里电影院银幕前的幕布在电影结束后要重新被拉回的原因。

新技术的最终用途通常是无法预料的，它远远超出我们最初想像力所及。电话已发展成无线通话，其数字通信网络可以让地球上任意两个人通过声音、文字和图像进行交流，而无论他们各自身居何处。1947 年，贝尔实验室发明了晶体管，很快人们就清楚地认识到这是个重大突破，但它最初的应用仅仅是针对旧的产品进行改善——晶体管取代真空管，这使得更小、更可靠的收音机和电脑出现。在当时，这已是相当重要和令人兴奋的变化了。晶体管最具革命性的应用直到最近才被发现。经过一段渐进的革新过程之后，人们开始设计出集成电路、微处理器、数字信号处理机和存储芯片。同样地，1970 年，为了改进桌上型计算器，微处理器发展起来了。它的最初亮相又一次被认为仅仅是替换“旧家伙”。因为当时的电子计算器仅仅是用于替换机械台式计算器，微处理器也很明显地成了替代螺

线管的候选技术(螺线管当时被用于某些工业控制领域,比如切换红绿灯交通信号灯。)然而,直到几年后,微处理器的真正威力才开始显现出来,而当时却没有人能预见到现代的个人电脑、手机、国际互联网、全球定位系统以及任何其他在今天看来是最基本的信息技术。

同样的道理,认为我们可以预言类大脑存储系统的革命性应用前景,那也是很不明智的。我完全相信并期待这样的智能机器将会在诸多方面改善我们的生活,对这一点我们深信不疑。但要预言这项技术若干年后的情形却是不可能的。只要回顾一下那些荒谬的预言家曾做出的充满信心的预言,你就能明白这一点。在20世纪50年代,有人预言到2000年时我们的地下室会有原子核反应堆;它能带我们去月球上度假。但只要我们能将这样的教训牢记在心,设想一下未来智能机器的样子还是有所裨益的。就其未来而言,我们至少可以得出一些广泛而又有益的结论。

我们能研制出智能机器吗?如果可以,那它们将会是什么样子呢?它们看上去是像我们在科幻片中常见的酷似人类的机器人呢,还是像现在个人电脑黑色或黄色的机箱盒,或者是别的什么样子?我们将会用它们来干什么呢?这项技术是不是很危险,会不会危害到人类并对我们的个人自由构成威胁?智能机器最明显的应用范围将是什么?我们有没有办法知道这些神奇的应用将会是什么?智能机器对我们生活的最终影响将会是什么呢?这些问题都很有意思。

我们能造出智能机器吗？

是的，我们能造出智能机器，但它可能不是你所想像的那个样子。我认为我们造出的智能机器可能不会如人一样行动，也不会以我们的方式与我们交流，虽然把智能机器做成人的样子是理所当然的事情。

大家熟悉的智能机器这个概念是从电影和小说上得来的。它们是外形像人的机器人，抑或可爱，抑或邪恶，抑或经常出错，它们与我们交流情感、思想，讨论问题，在科幻小说的情节发展中扮演着重要角色。上百年来，科幻小说使人们深信，机器人或类人的机器是我们未来生活不可避免并令人期待的一部分。几代人都是伴随着《惑星历险》(Forbidden Planet)中的罗比(Robbie)、《星球大战》(Star Wars)中的 R2D2 和 C3PO 以及《星际迷航》(Star Trek)中的科学上尉(Lieutenant Commander Data)的形象成长起来的。甚至于《2001：太空漫游》中的哈尔(HAL)，虽然没有身躯，但也非常像人，它被设计成人类漫长太空之旅中的同伴，同时还是由程序控制的副驾驶员。能被人类所控制的机器人是可实现的，诸如智慧型汽车(smart cars)、自动深海探测器、能自动导向的真空吸尘器或割草机之类，它们有朝一日会变得越来越普及。但人形的机器人、科学上尉和 C3PO 等等，它们在很长一段时间内仍将是虚构的。这有几个方面的原因。

首先，人的心智不但产生于新大脑皮层，还需要古脑的情感体系和人体的复杂结构。作为人，你不光需要大脑皮层，还需要具备人的一切生物学机体。一个智能机器要能像人一样就

一切问题进行谈话(如通过图灵测试),它就必须要具有人的大部分经历和情感,并要过着像人一样的生活。智能机器将拥有与人类大脑皮层功能一样的器官和一组感官系统,其他倒是可选的。看着人形的智能机器蹒跚着来回走动可能是件很有趣的事,但它们却远远没有人一样的心智,除非我们让它们具有人的情感系统和经历,但那将是非常困难的。而且对我来说,这样做也没有多大意义。

其次,制造和维护类人机器人需要巨大费用和精力,这样的话,它们便很难成为实用的机器了。一个机器人管家比起真人助手来,费用更高且帮助甚少。虽说机器人可能会有“智能”,但它却不能像真人助手那样与人相处融洽、和谐,因为真人助手毕竟是我们的同类。

蒸汽机和数字计算机都引发了人们对机器人视觉的研究,但这方面的研究却一直没有完成。同样地,当我们想到要制造智能机器时,许多人很自然地就再次想到了类人机器人,但这却似乎不可能发生。机器人是源于工业革命的一个概念,并在小说中得到升华。在研究真正意义上的智能机器这个问题上,我们不能向虚构的“机器人”寻求灵感。

如果智能机器并不是会走会说的机器人,那它们会是什么样子呢?我们发现,如果在进化过程中给我们的感官连接上一个分层的存储系统,那么这个存储系统就会建立起一个关于世界的模型,并以此预测未来。我们从自然界那里认识到,我们应该沿着与此相同的路线制造智能机器。那么,下面就是制造智能机器的方法。

首先,我们需要一组感觉器官,通过它们从世界提取模

式。我们的智能机器将有一组不同于人类的感觉器官，这些感觉器官甚至可能是“存在”于一个有别于我们所熟悉的世界里的。因此，不必一定要认为它要有一双眼球和一对耳朵。

其次，跟这些感官相连的是一个以大脑皮层工作原理而运作的分层存储系统。接下来就是要像教育自己的孩子一样去训练这个存储系统。经过反复的训练，这个智能机器就会通过自己的感官构建起一个它自己的关于世界的模型。在此过程中，我们不需要，也没有必要去编程序、建立规则来描写这个世界，也不需要数据库、知识库或任何高级的概念。这些都是人工智能领域中最令人头痛的部分。

智能机器必须通过观察它周围的世界来进行学习，必要时也包括接受训练者的输入。一旦我们的智能机器建立起一个关于世界的模型，它就能依据过去的经历做出类推，从而对未来事件做出预测，为新问题提供解决方案，并且还可以帮助我们得到某些知识。我们的智能机器可能将作为部件安装于飞机或汽车之上，或放置于计算机机房的架子上。与人不同，智能机器的记忆系统可能与其感官(或“躯体”，如果有的话)相距甚远，比如说，一个智能安全系统的感觉传感器可能分布于一家工厂或某个城市，而与这些传感器相连的分层记忆系统却被锁在一幢大楼的地下室，而人的大脑总要与身体相伴相随。因此，智能机器的物理形态可以是多样的。

没有理由要求智能机器一定要有人的形象，能像人一样行动，具有与人一样的感觉。智能机器之所以有智能，是因为它可以通过一个分层记忆系统来理解它的世界并与之交互，可以

如你我一样思考自己的世界。正如我们所看到的，它的思维和行为可能与我们的完全不同。但，即便是这样，它仍然是有智能的。智能是通过分层记忆系统的预测能力进行衡量的，而不是看它是否具有与人一样的行为。

让我们将注意力转移到制造智能机器所面临的最大挑战——建立记忆。要制造智能机器，我们需要建立一个有着分层结构并如大脑皮层一样工作的大型记忆系统。在这个问题上，我们将要面临的是记忆容量和连通性方面的挑战。

容量是第一个问题。我们说大脑皮层有 32 万亿个突触。如果我们只用两个比特表示一个突触(这样，每个突触便有 4 种可能的值)，而每个字节有 8 个比特，结果，我们就需要大约 8 万亿字节的记忆容量。现在个人电脑上的一个硬盘能达到 100 吉(1 吉等于 1 024 兆)字节的容量，也就是说，我们需要大约 80 个这么大的硬盘才能有相当于人的大脑皮层的记忆容量。(不必计较精确的数值，因为这都是粗略的估计。)问题在于，这么大的记忆容量只有在实验室才能实现。(当然，这等同于在 10 年前也许是不可能的)，且仍没能小到可以放进你的口袋或电烤炉里。所幸的是，我们不必重建整个人的大脑皮层那么大的记忆容量，因为很少的记忆就完全能满足我们的需要了。

我们的智能机器会需要大容量的记忆。也许一开始我们将用到硬盘或者光盘，但最终我们要用硅晶片来制造它们。硅晶片尺寸比较小，功耗低，并且是蚀刻的，因此它被制造成有足够的容量并用于智能机器只是一个时间的问题。实际上，比起传统的计算机存储器来，智能存储器有一个优势。半导体工业的经济情况取决于芯片的废品率。芯片的良品率就叫做收益

率。它决定了某种芯片设计是否可以投入生产并能否创造利润。因为随着芯片尺寸的增大，废品的几率也随之增加，所以现在大多数的芯片只有一枚小邮票那样大小。芯片业在单个芯片上增加记忆容量不是通过增大芯片面积，而是通过使芯片上的记忆单元更小而实现的。

但是智能记忆芯片却天生具有容错能力。请记住，你大脑中并没有哪一部分独立地记录着不可或缺的数据项目。你的大脑每天都会有成千上万个神经元死去，但你的心智能力在整个成年期却以非常缓慢的速度在衰退。智能记忆芯片会以大脑皮层同样的原理工作，因此即使记忆单元中有一部分损坏了，芯片仍然可用并仍具有商业价值。很可能的是，类人脑记忆的固有容错能力可以让芯片设计者设计出比现有的计算机存储芯片大得多、厚得多的芯片来，结果是，我们以比当前发展更快的速度将大脑植入硅晶片。

连通性是我们克服的第二个问题。人脑有大量的皮层下白质。我们在前面已经看到，这些白质是由上百万个的轴突组成，这些轴突在薄薄的皮层下朝着各个方向延伸，从而连接着皮层各层次中的不同区域。皮层中的单个细胞会与 5 000 ~ 10 000 个其他细胞相连。这种大量的平行布线运用传统的硅生产技术是难以甚至是不可能做到的。硅晶片是通过将几层硅金属叠压而制造出来的，每一层都隔有绝缘层。（这个生产过程与大脑皮层中的层次没有相似之处。）每层金属中都含有芯片的“线路”，因为各层的“线路”不可能交会，所以连接的总数是有限的。这对于类人脑的记忆系统毫无意义，因为必须要有上百万的连接才行。硅晶片与白质并不很相似。要解决这个问题

题还需要进行大量的实验和工程革新，但我们已知道了解决之道的的基本方面。

电线传递信号的速度远快于神经元的轴突。芯片上的单条线路可以进行分配，因而可用于许多不同的连接，但人脑中每个轴突仅属于一个神经元。

电话系统就是一个实例。如果我们在每两部电话间都连上电话线，那么地球表面将会淹没在由电线交织而成的丛林之下了。实际上我们是让所有的电话共享一组数量相对较小的大容量电话线。只要每条线的容量远大于传递单个通话所需的容量，这种方法就能奏效。电话系统就满足这一条件：单根光缆就能同时传送上百万个通话。

人脑中所有细胞间都有各自专有的轴突相连，使细胞得以彼此沟通。但我们研制的智能机器却更像电话系统，共享着连接。不管你相信与否，有一些科学家们已经花了数年时间在考虑如何解决脑芯片连通性的问题。给大脑皮层动手术仍是件令人不可思议的事，但研究人员相信我们迟早会解决这一难题的，到那时，我们将不得不面对连通性的问题。在此，我们不必回顾那些曾经尝试过的多种方法了，我们只需说连通性是我们研制智能机器过程中面临的最大的技术难题就可以了。我们终究会解决这个问题。一旦技术难题得以解决，就没有什么根本性的问题能阻挡我们制造真正的智能系统了。诚然，为了使智能系统更小、成本更低、功耗更低，我们还有许多问题要解决，没有什么可以阻挡我们。计算机从最初的房间大小到现在能放进口袋里一共花了 50 年的时间。但由于现在我们处于更高的技术起点上，向智能机器过渡将会快得多。

我们该不该制造智能机器呢？

在 21 世纪里，智能机器将从科幻小说中走出来变为现实。在这之前，我们应该考虑一下伦理问题，即智能机器潜在的危险是否会超过它可能带来的好处。

能思考、能按自己意愿行动的机器的未来景象已困扰人们很长时间了，这是可以理解的。新知识和新技术在出现之初总会使人们感到恐慌。人们具有创造性思维，因而会想像出可怕的情形来，比如新技术会取代我们的身体，使我们不再有用，使我们的人生不再有价值。但历史已证明，这些可怕的设想从未如我们料想的那样出现过。

工业革命之初，我们惧怕电[还记得弗兰肯斯坦(Frankenstein)其人吗?]和蒸汽机。机器自己能提供能量并能以复杂方式移动，这似乎很神奇，但同时也可能很危险。然而，电和蒸汽机不再让我们感到奇怪和危险，它们就像空气和水一样已成为了我们环境的一部分。

当信息革命开始时，我们很快就对计算机产生了恐惧感。有数不清的关于强大的计算机和计算机网络的科幻小说，在这些小说里，计算机一旦有了自我意识，就会与自己有生命的主人为敌。但如今电脑已成为我们日常生活的一部分，那种担心看起来似乎太荒谬了。你家的电脑，互联网与收款机也一样，几乎不可能变得有知觉。

当然，任何技术都有可能用于正义或邪恶的目的，但其中一些可能更容易被误用而给别人带来灾难。核能不论是用于核弹头还是发电厂都是危险的，因为一次小小的事故或一次误操

作就有可能危及数百万人的生命。虽然原子能很有用，但其替代能源也已经有了。

运输技术可以采取坦克和战斗机的形式，也可以采用汽车和民用飞机的方法，但一次事故或误操作会给许多人造成伤害。事实证明，车辆已成为现代生活不可或缺的一部分，同时它们还不具有核能那样的危险；飞机误操作所造成的损失远远小于核能的误操作。有许多技术几乎绝对是有益的。比如说电话，它能让人们保持联系，给我们带来的正面效果毫无疑问要多于它所带来的负面效应。电和公共卫生学也是一样。依我看来，智能机器将会是我们所开发的技术中危险最小而益处最多的发明。

但仍然有像 Sun Microsystems 公司的共同创始人比尔·乔伊 (Bill Joy) 那样的思想家，他们担心我们有可能研制出微型智能机器人，这些机器人有朝一日能摆脱我们的控制，遍布地球并根据自己的计划一点点地重建这个世界。这幅图景使我想起了《魔法师的徒弟》(The Sorcerer's Apprentice) 中那魔术般活动着的扫帚柄，它们以碎片重建自己，并不厌其烦地给人们带来灾祸。类似地，一些对人工智能持乐观态度的人做出了这一技术能延长生命的预言，这使我们感到不安。比如说，雷·柯兹威尔 (Ray Kurzweil) 曾谈到，有那么一天，微型机器人钻进你的大脑，记录下每一个突触以及每一次连接，然后将所有的这些信息报告给一台巨型计算机，这台巨型计算机就能重新调整自己的配置，进而成为另一个你。你将变成一个自己的“软件”版本，并且几乎是不朽的。人们对机器智能的这两种担心——智能机器疯狂运作带来混乱以及把你的大脑上传到电脑的设想

——好像还将不断涌现。

制造智能机器与制造自我复制机器其实并不是一码事，两者之间没有任何的逻辑联系。人脑和电脑都不能直接进行自我复制，与大脑类似的记忆系统也是一样。而智能机器的优越性之一就是我们可以对它们进行大规模生产，当然这种大规模生产并不像细菌和病毒那样进行自我复制。自我复制不需要智能，智能也不需要自我复制。

此外，我很是怀疑我们能否将心智复制到机器中去。据我所知，目前还没有任何现实或想像中的方法能把成其为“你”的那些上万亿的信息细节记录下来。如果要复制你，我们要将你的神经系统和身体的信息都进行记录并进行再创造，而不仅仅是你的大脑皮层。我们还要进一步了解它们都是如何工作的。当然很可能有那么一天，我们将有能力完成这一切，但面临的挑战要远远大于理解大脑皮层的工作原理。弄清新皮层的算法并将其从人脑植入机器是一回事，将有生命的大脑中无穷的操作细节扫描并复制到机器中，却是完全不同的另一回事。

除了自我复制和心智复制以外，对于智能机器人们还有另外一些忧虑。智能机器会如原子弹那样对人类构成威胁吗？它们的出现会导致小集团或邪恶的个人进行独裁吗？智能机器会如《终结者》和《黑客帝国》中那些不可控制的机器一样作恶并与我们作对吗？这些问题的答案是否定的。作为信息设备，类似大脑的记忆系统将是我們开发出来的用途最大的技术之一。但就像汽车和电脑一样，它们仅仅是工具而已，具有智能并不意味着它们将

有特殊的能力来破坏我们的物质世界或控制人类。正如我们不会将世界上的核武库置于某个人或某台电脑的控制之下一样，对于智能机器我们也要谨慎，不要太过于依赖它们，因为它们像所有其他的技术一样有可能会带来破坏。

这样，我们又面临着另一个可怕的问题。有人认为机器具有智能基本上就等于说它具有了人的智力。他们担心智能机器将会因为“被奴役”而感到愤怒，因为人类憎恨被奴役。他们担心智能机器会试图统治世界，因为历史上曾有智者这么做过。但这样的担忧是建立在一个错误的类推之上的。人们之所以有这样的担忧是因为他们将智能——新皮层的算法，与古脑的情感因素——诸如恐惧、多疑和欲望等，归并起来了。智能机器是不具备这些能力的。它们不会有野心，也不会渴望财富，寻求社会认同以及性满足。它们没有食欲、嗜好，也不会出现情绪不稳定的情况。智能机器不会有任何类似人类情感的东西，除非我们刻意把它们设计成那样。智能机器最强大的应用就是那些人类智力有困难的地方，感觉器官不能及的领域和那些单调乏味的工作。通常，这样的活动几乎不涉及情感。

智能机器应用的范围从简单的单一应用系统到强大的超人智能系统无所不包，但除非我们自己要把它们设计成人的模样，否则它们是不会像人类的。也许有一天我们会制定规范来限制智能机器的应用范围，但那一天还很遥远。当那一天真的到来时，它所涉及的伦理问题，将比现在诸如围绕基因和核技术所引发的道德问题要相对容易解决一些。

为什么要制造智能机器？

现在我们来看看这个问题——智能机器能做什么？

经常有人要我就移动计算的未来发表看法。会议组织者总是会要我描述一下 5 年或 20 年之后的掌上电脑或移动电话将是什么样子，他们想听听我对未来的预计，但我办不到。我总是要避免回答那些涉及观点的问题。为了表明我的立场，有一次我戴着魔术帽，拿着玻璃球走上讲台。我解释道，任何人都不可可能看清未来的每个细节。玻璃球是虚幻的，任何装着想确切知道未来几年将发生什么的人注定是要失败的。反过来，我们所能做的就是把握大的趋势。如果你领会了总的思路，那么不管细节明朗化后它将如何发展，你都能成功地把握住它。

预言技术发展趋势最有名的例子就是摩尔定律。戈登·摩尔(Gordon Moore)准确地预言了一块硅晶片上能容纳的电路元件数目每一年半就会翻一番。当时，摩尔没有说这些芯片是否是存储芯片、或者是中央处理单元，或者什么其他的东西，并没有说这些芯片将用于何种产品，也没有预测这些芯片将放置在塑料或陶瓷里，或黏附在电路板上，更没提及任何制造芯片的加工工艺。他仅仅是把握了大体的发展趋势，然而他是完全正确的。

现在我们无法预言智能机器的最终用途。我们没有办法准确地了解到那些细节。如果我或者任何其他人详细地预言了这些机器的用途，那不可避免地将会被证明是错误的。尽管如此，我们能做的不仅仅是耸耸肩而已，有两条思路或许可以帮助我们。其一就是想像一下这种类人脑的记忆系统最短期的用

途——尝试一些显而易见、比较缺乏趣味性的应用；其二就是考虑一下长期的趋势，就像摩尔定律一样，这有助于我们想像可能成为未来生活一部分的智能机器的应用。

让我们先从短期的应用着手，这些应用都是显而易见的。这就类似于用晶体管替换收音机里的电子管，或者用微处理器来制造计算器。我们可以先涉足那些人工智能尝试要解决却又没有完全解决的领域，比如说，语音识别、机器视觉和智慧型汽车，等等。

如果你曾尝试过用语音识别软件将文本输入个人电脑，你就会知道这样做是多么的愚笨。就像塞尔的“中文屋”实验，计算机是不会明白你所说的话的。经过几次尝试，我对它彻底失望了。如果房间里有任何噪音，即便是铅笔掉落到地上的声音或者别人对我说话，都会导致屏幕上出现多余的字词，它的错误率很高。软件识别出我所说的词没有任何意义，这样的情况常常出现。“Remember to tell Mary that the bog is ready to be piqued up. (记住摔倒玛丽，沼泽已做好发怒的准备!)”小孩都能知道这是错的，而计算机却不晓得。同样，所谓的“自然语言接口”已成为计算机科学家多年来研究的目标，其内容就是通过人的日常语言告诉计算机或其他的设备你需要什么，并让它们完成相应的工作。在 PDA 上，你可能会写道：“把我女儿星期天的篮球比赛改到今天上午 10 点。”传统的人工智能技术不太可能做好这项工作。即便是计算机能识别每一个字，要完成这项工作却需要它先知道你女儿在哪儿上学，也需要知道你

大概指的是下个星期天，还需要知道你所说的“篮球比赛”到底指的是什么，因为它有可能只被当作是“门罗队对圣乔队”。也许你要计算机听广播并要求它判断某个产品的名字是否被提到，而广播员却只描述了产品，没有提到该产品的名字。你我都能知道播音员说的是什么，但计算机却判断不了。

诸如此类的应用都要求机器能听懂人类的口头语言，但计算机却无法完成这样的任务，因为它们听不懂我们所说的话。计算机通过声音将听觉的模式与单词模板进行匹配，根本不知道这些单词的意思。假设你学着能辨出一种外语的每个单词的发音，但不知道它们的意思，然后我要你记录下用那种语言进行的一段会话。在会话进行过程中，你根本不知道它是关于什么的，但你却尽力要把一个个单词辨识出来。结果呢，单词的发音有交叠并互相干涉，还有噪音也使你没能听到一些声音片段，你会发现要把一个个单词进行分离并识别出来有多么的困难。这正是开发语音识别软件目前所面临的障碍。工程师发现通过运用词过渡概率(probabilities of word transitions)，多少可以提高软件的准确率。比如说，他们利用语法规则来判定同音(同形)异义词等。这是预测的非常简单的形式，但还是不具有智能。今天的语音识别软件在非常受限的条件下可以取得一些成功，因为这些条件限制你在一定的时间里只能说出有限的单词。

人之所以能顺利地完成任务，是因为我们的大脑皮层不但能理解单词，还能理解它们所在的句子和上下文。我们在期待思想、短语和单个的词。我们大脑皮层有关世界的模型自动地在完成着这些任务。

因此，我们可以预期类皮层的记忆系统会将脆弱的语音识别变成强健的语音理解。不用借助词过渡概率进行编程，我们的分层记忆可以追踪音调、单词、短语和思想，并利用它们来理解所说的话。和人一样，这样的智能机器可以区分不同的话语事件，比如，是你和朋友在房间里进行的讨论，还是一次电话交谈，还是编书时的编辑指令。研制这样的机器不容易。要完全理解人类语言，机器必须经历人所经历的，学习人所学习的。因此，虽然研制能像你我一样可以理解人类语言的智能机器还需要许多年的时间，可在短期内，我们可以利用类皮层的记忆来提高现有语音识别系统的性能。视觉方面也有一些应用是人工智能无法完成而真正的智能系统却可以应付的。今天还没有什么机器能欣赏自然风光——我们眼前的世界——或者照相机拍下的景象，并能把它描述下来。现在已有少数能在人们规定的领域里成功工作的机器视觉应用，例如在电路印刷版上通过视觉定位芯片或比照数据库匹配五官特征，但目前还无法让计算机识别各种物体，或者在更广泛的意义上分析所看到的景象。环顾一下房间并找个位置坐下，对你来说没有什么困难，但不要期望计算机也同样能做到。假设看着接有安全监视器的电视屏幕，你能否区分一位手拿礼物敲门的人与一个企图用起货钩撞门的家伙呢？当然可以。但两者之间的差别是目前软件所不能胜任的。于是我们必须雇人日夜监控监视设备的屏幕，以便发现任何可疑之处。人要长时间保持警惕是很困难的，但智能机器却可以不知疲倦地完成这项任务。许多需要敏锐视觉完成的工作同样需要对复杂场景的理解，智能机器是处理这些任务的唯一途径。

最后，我们来看一下运输。汽车现在越来越复杂精密了，它们装有全球定位系统，可以帮你确定从 A 点到 B 点之间的路线；装有天色变暗时自动开灯的装置，装有控制安全气囊的过荷传感器；装有“接近觉传感器”，以防你倒车时撞到什么东西。现在，甚至还有能在专用公路或在理想条件下自动驾驶的汽车，虽然它们还没有商品化。但要在各种公路上和任何交通条件下安全有效地驾驶，还需要更多的传感器和反馈控制电路。要成为一名合格的驾驶员，你必须了解车辆、其他的驾驶员、汽车的工作原理、信号灯和许多其他东西。你必须能看懂道路危险标志，能注意到另一辆车是否在危险驾驶。你还必须能看到另一辆车上的转向信号灯，并预期它很可能要换道行驶。而如果信号灯亮了好几分钟，你应该能意识到可能是那个司机并不知道信号灯亮着，因而也不一定是打算换道行驶。当前方有烟雾冒起时，你要能意识到可能发生了事故，就应该减速行驶。当看到球滚过马路时，你会不假思索地判断出很可能会有孩子跑出来捡球，而凭直觉将车刹住。

如果说我们要造一辆真正意义上的智慧型汽车，首先要做的就是选出一组传感器，让我们的智慧型汽车能感知周围的情况。要在车的前后装上一个或多个能“看”的摄像头，能“听”的拾音器，还需要装上雷达系统或超声传感器，以便在明亮或黑暗条件下准确地判断其他物体的距离和速度。问题是，我们其实不必仅仅依赖或局限于人所使用的感官。脑皮层算法是灵活的，只要我们以适当的方式设计我们的分层记忆系统，不管我们装了何种传感器，它都应该能很好地工作。理论上讲，我们的汽车能比我们更好地感知交通状况，因为它的各

种感觉都可以用来完成任务。我们把这些传感器连接到一个足够大的分层记忆系统上，通过将这种智慧型汽车置于真实的条件下来训练它的记忆，它将像人一样学习并构建世界的模型——当然是在有限的范围内。（比如说，它必须了解道路而不必知道电梯或飞机等。）它的记忆将会记住交通和道路的分层结构，这样，它就能在有着穿梭行驶的汽车、道路标志、障碍物和交叉路口等的真实世界里明白并预测将要发生什么。汽车工程师设计这样的记忆系统，以便汽车能自动驾驶或在我们开车时能提供监控、给出建议，或在极端条件下接管汽车，就像一名你不会讨厌的后座驾驶员。记忆一旦训练好，汽车将可以理解并知道如何处理各种状况。

在汽车走下装配线时，汽车工程师可以使其永久地固化记忆，这样，每辆车都会有同样的表现。我们还可以将汽车的记忆设计成售后可继续学习型，因为车上装的是电脑而不是人脑，所以如果条件允许的话，我们可以将记忆重新配置，让它成为升级版本。

我并不是说我们肯定会制造出智慧型汽车或者能理解语言、有视觉的机器，但这是我们能研究和开发的装置中的极好例子，并且这些机器也有可能被制造出来。

就我个人而言，我对智能机器的那些较明显的应用并不怎么感兴趣。在我看来，一项新技术真正的好处以及它能带给我们的兴奋在于它前所未有的用途。智能机器到底会在什么方面带给我们惊奇呢？随着时间的推移，又会出现哪些神奇的能力

呢?我确信,正如晶体管和微处理器那样,分层记忆将会在许多方面改善我们的生活,可它又是怎样改善我们生活的呢?为了管窥智能机器的未来,我们可以考虑一下这项技术能很好扩展的那些方面。也就是说,看看智能机器的哪些方面成本会越来越低,速度会越来越快,或体积会越来越小;看看哪些能以指数级速度增长的事物会迅速超出我们的想像,并很可能在未来技术的根本性变革中起到重要作用。

多年来,能以指数级速度发展的技术有硅晶片、硬盘、DNA 排序技术和纤维光学等等,这些飞速发展的技术已经成为许多新产品和事业的基础。软件也以另一种方式在迅速发展,一项能满足人们需求的程序一旦被开发出来,就能无成本地被无限复制。

相反,另外一些技术,比如电池、发动机和传统机器人等,它们的发展都太缓慢了,虽然人们付出了很多努力并且也有了稳定的进展,可今天造出的机器手臂与 10 年前的相比并没有多大改观。机器人的发展是缓慢的,这与以指数曲线发展的芯片设计和蓬勃发展的软件业截然不同。1985 年用 100 万美元造的机器手臂在今天不会变得只要 10 美元就能造出,并且其功能还能再强大上千倍。同样地,今天的电池也并不比 10 年前的有多大改善,也许有两到三倍的提高,但却没有成千上万倍的提高。如果电池的容量发展速度也像硬盘一样,那么移动电话和其他一些电子设备就不需要费用了,并且那些开上 1 000 千米以上才需要充电的轻型电动车也会很普遍。因此,我们应该好好想想类人脑的记忆系统在哪些方面会以超过人脑的速度迅速发展,而这些方面将会表明这项技术的最终归属。

我认为智能机器在 4 个方面会超过我们人脑的能力：速度、容量、可复制性和感觉系统。

速度

当神经元以 $1/10^3$ 秒的速度工作时，硅正以 $1/10^{10}$ 秒的速度运行（而且这一速度还在不断加快）。这是 100 万倍的差异，或者说有 6 个数量级的差别。人脑思维和基于硅片的思维之间速度上的差异将会给我们带来惊人的结果，那便是，智能机器能够以快于人脑 100 万倍的速度思考。这样的脑子可以在几分钟内读完整个图书馆的书或者研究大量的资料，而同样的任务你我都需要几年的工夫才能完成，并且它们的理解与人无异。这其实没有什么不可思议的，因为人脑的进化有两个与时间有关的限制，其一是细胞工作的速度，其二就是世界变化的速度。对人脑来说，如果它周围的世界以很慢的速度变化发展，而它自己却以快上 100 万倍的速度工作的话，那将毫无意义。但脑皮层算法并不意味着它必须以很慢的速度工作。如果一台智能机器与人交谈或交互，它就必须慢下来并以人脑的速度运行；如果它读书时要翻页，这也将限制它的阅读速度；但如果当它与电子世界交互时，它就可以运行得非常快——两台智能机器间交谈的速度可以比两个人交谈的速度快上 100 万倍。想像一下智能机器以比人快上 100 万倍的速度解决数学或科学难题的过程。它在 10 秒钟内对一个问题所作的思考，你可能需要花上 1 个月的时间。这种不知疲倦、不知厌烦，并能以闪电般速度进行思维的机器，将会在我们所想像不到的方面为我们造福。

容量

不管人脑皮层的容量有多大，智能机器的记忆容量都可以远远地超过它。人脑的大小受到一些生理因素的制约，其中包括：婴儿头颅的大小取决于母亲骨盆的直径，大脑工作时所需很高的代谢（你的大脑虽然只占你体重的2%，但它却要消耗你20%的氧气），以及神经元较慢的工作速度。但我们制造的智能记忆系统可以是任何大小的，而且在设计的细节上我们可以做精心的调整。相比之下，几十年后人类新皮层的容量将会显得有些捉襟见肘了。

制造智能机器的时候，我们可以从多方面扩大它们的记忆容量。例如：在体系中增加层次将会导致更深入的理解——具备理解更高层次模式的能力；扩大各区域的容量将会使机器能记住更多的细节，有更敏锐的观察力，正如盲人有更敏锐、更精确的触觉和听觉一样；还可以扩展感觉系统的层次，让它能构建更好的世界模型。我马上就会对此进行讨论。

看智能记忆系统的规模是否有上限及其扩展方向，是非常有趣的一件事情。可以想像，一个装置可能因为太复杂而不实用，或者当它接近某个理论极限时就可能失效。也许人脑的大小已接近理论极限了，但我却认为这不太可能。人脑是在最近的进化时期才增大的，也没有什么迹象表明我们的大脑正处于某种比较稳定的最大尺寸。但不论智能记忆系统的最大容量会有多大，人脑几乎都是无法企及的，甚至是丝毫无法接近的。

如果想知道这些系统能有何作为，我们不妨看看已知的人类行为的极限。爱因斯坦毫无疑问是绝顶聪明的，但他的大脑



仍然是人的大脑。我们可以假定他非凡的智力很大程度上是因为他的大脑与典型的人脑之间存在的物理上的差异。爱因斯坦这样的人之所以少，是因为我们的基因不会经常产生他那样的大脑。然而，当我们用硅设计大脑时，我们可以以任何可能的方式制造它。它们可以都具有爱因斯坦的高级思维能力，甚至更加聪明。在另一极端，“奇人”可以使我们认识到智能待开发的其他潜能。“奇人”是那些心智发展不全的人，但他们却在某些方面表现出非凡的能力，比如具有近乎照相机般的记忆，或者具有飞速计算复杂算术题的能力。他们的大脑，虽然不是典型的大脑，但终究还是大脑，同样以大脑皮层的通用算法工作。如果一个非典型的大脑能有惊人的记忆能力，那么，在理论上我们就可以将这些能力赋予人造大脑。这些人类非凡能力的极端例子不但表明了什么是可能再创造的，而且还表明了到什么方向上我们也许能超越最佳的人类行为表现。

可复制性

每个人脑都必须从头开始成长并接受训练，这一过程要耗去人类几十年的时间。每个人都要学会如何协调肢体和肌肉组织，要学会如何平衡、如何移动，还要掌握大量事物、动物和他人的基本属性，要记住事物的名称和掌握语言的结构，还要学习家庭和社会的规范。在掌握了这些基本技能和知识之后，便开始了年复一年的正规学校教育。

每个人都要在人生相同的学习道路上艰苦跋涉，即使这些过程已被无数人一次又一次地重复过，这样做的目的就是在人脑皮层中建立起世界的模型。

智能机器不必经历这种漫长的学习过程，因为芯片以及其他的存储器都可以被不断地复制，其中的内容也可以被轻易地转录。从这种意义上讲，智能机器可以如软件般复制。一旦我们对一个训练好了的原型系统感到满意，就可以随心所欲地将其复制。为完善智慧型汽车所进行的芯片设计、硬件配置训练和反复试验需要花上许多年的时间，但一旦完成了最终产品，就可以投入批量生产。正如前面提到过的，我们可以选择是否让这些复制品继续学习。

在某些应用领域，我们可能要限制智能机器使其按照经实验通过的方式运行。一旦智慧型汽车掌握了我们希望它掌握的东西，我们就不希望它继续学习，因为那样可能使它养成不良的习惯或者坚持它自以为正确而实则错误的类推。我们可能希望所有相似的汽车都有相似的表现，但对于其他的一些应用，我们要求这种类大脑的记忆系统具有继续学习的能力。例如，设计成解数学题的智能机器必须具备从经验中学习的能力、具备依旧知识洞察新问题的能力，此外还要相当灵活，能应付各种开放性问题。

我们可以像分享软件组件一样分享智能机器的学习组件。某种特定设计的智能机器可以通过新的连接重新编程，从而具有不同的行为，就好比我可以下载一些新的神经连接到你的大脑中，将你即刻从讲英语的人变为讲法语的人，或者从政治学教授变成音乐研究者一样。

人们可以交换、重复利用彼此的工作。比如说我开发并训练了一种有超级视觉系统的机器，而别人却开发并训练了一种有超级听觉的机器，通过适当的设计，我们可以将这两个系统

组合起来，从而大大节省了时间。人是不可能以这样的方式来分享专业知识和技能的。智能机器制造业可以采取计算机工业发展的路线来发展自己，不同的团队从事面向不同专业和能力的智能机器的训练，通过购买和交换彼此完成的记忆配置，改编一个智能机器就如同运行一种新的电脑游戏或安装一种新的软件一样容易。

感觉系统

人类有少数几种感觉，这些感觉根植于我们的基因、身体和大脑皮层下面的神经网络里，我们无法改变它们。有时我们可以利用技术扩大感觉的范围，比如可以使用夜视镜、雷达或者哈勃空间望远镜等。这些高科技仪器设备并不是什么新的感知方式，它们只是巧妙地将数据进行转换，将我们不能感知的信息转换成可理解的视觉或听觉的显示。

人脑有着惊人的灵活性，我们看着雷达屏幕就可以明白它表示什么。许多种动物都有着真正不同的感官，比如蝙蝠和海豚有回声定位系统，蜜蜂能看到偏振光和紫外光，某些鱼能感应电场。

智能机器可以通过任何自然界已有的感觉方式，还可以通过由人全新设计的感知方式来感知世界。声纳、雷达、红外线观测仪都是我们可以让智能机器具有的非人类感觉的明显例子。它们仅仅是个开始。

更有趣的是智能机器能体验到真正奇异的感觉世界。正如我们看到的，新皮层算法从根本上来讲是用来发现世界模式的，而不仅仅倾向于这些模式的基本物理特性。只要输入到皮层的

信息不是随机的，并有一定的丰富性和统计特征，智能系统就能对它们形成恒定记忆并进行预测。没有理由要求这些输入模式一定要与动物的感觉类似，也不需要它们完全由现实世界派生而来。我以为，智能机器革命性的用途就在于这些奇异的感官领域。例如，我们可以设计一种跨越地球的感觉系统，可以在陆地上遍布天气传感器，把地面分隔成大约 70 千米的小块，它们就像是视网膜上的细胞，在任一时刻，两个相邻传感器的活动都有着高度的相关性，就如同视网膜上两个相邻的细胞那样。风暴、锋面等巨大天气对象会随时间而移动变化，就像视觉对象会随时间移动变化一样。传感器将这些天气连接到一个巨大的类皮层记忆系统，我们就可以让它学着预测天气，就像你我学着识别视觉对象并预测它们如何随时间而移动那样。这个智能系统将能识别局部天气模式、大规模天气模式，以及那些以几十年、几年或几小时为发生间隔的模式。通过在某些地区放置更密集的传感器，我们就可以产生类似视网膜上视觉凹的结构，从而使我们的智能机器能理解并预测更细微的天气变化。我们的天气智能系统可以思考和掌握全球的天气系统，就像你我能思考和理解事物和人一样。今天，气象学家正致力于建造类似这样的系统，他们收集不同地区的记录，并利用巨型计算机来模拟天气，然后预报未来的天气。但这种工作方式与智能机器工作的方式有着本质区别，却与电脑博弈类似，因为它和电脑博弈一样并不理解自己在做什么。而智能机器的工作方式却与人下棋是一样的，都在做周详的考虑并能理解自己所做的一切。

智能天气系统尤其要能发现人发现不了的模式。厄尔尼诺

现象是在 20 世纪 60 年代才发现的，而我们的智能天气系统可以发现更多类似厄尔尼诺现象的模式，也可以学会比人更好地预报龙卷风和季风。要将大量的天气资料变成人容易理解的形式不是件容易的事。相反，我们的天气智能系统却能直接地感知和思考天气现象。

其他大型的分布式感觉系统能让我们建成可以理解和预测动物迁徙、人口变化和疾病传播的智能机器。设想一下，我们可以在一个国家的电力网上分布传感器，与之相连的智能机器能观测到耗电量的消长，就像你我能观察到道路上交通的变化或机场里人员的流动一样。通过反复地接触这些环境，人们学会了预测这些模式，只要问一下开车上下班的职员或机场的保卫人员就可以明白这一点。同样，智能电网监控系统能比我们更好地预测对能源的需求，能预测到可能导致电力中断的危险状况。

我们还可以将天气传感器和人口分布传感器结合起来，以预测政治动乱、饥荒或疾病暴发。智能机器能像一名睿智的外交官，可在减少冲突和人类痛苦方面扮演重要角色。也许你会认为智能机器必须要有情感才能预测与人类行为有关的模式，而我却不这么看。我们不是生来就属于某种文化，生来就具有某种价值观，或生来就信仰某种宗教的，它们都是后天习得的。正如我可以学会理解与我有不同价值观的人的动机，智能机器也可以理解人类的动机和情感，即便它自己并不具备这些动机和情感。

我们可以研制出能提取微小物体的感官。传感器要能用细胞或大分子来表征模式，这在理论上是可能的。例如，我们现

在面临的一个重要挑战就是如何通过构成蛋白质的氨基酸序列来预测蛋白质分子的形状。能够预测蛋白质将如何折叠和交互，就能促进发展医学的进步并能治疗多种疾病。工程师和科学家已经建立了蛋白质的立体可视化模型，目的就在于要预测这种复杂分子的表现。即使我们已尽己所能，事实证明这项工作是非常困难的。采用装配着针对这一问题而特别设计的感觉装置的超智能机器，这一难题或许会迎刃而解。这听起来好像有点令人难以置信，但请记住，如果最终解决了这一问题，我们是不会感到惊讶的。这一问题之所以难以解决，主要还是因为在我们自身的感觉系统与想要解决的问题之间存在着很大的差距。智能机器有定制的感官，还有比人脑容量大的记忆，它们完全可以解决我们所不能解决的问题。

通过合适的感觉和稍加调整的皮层记忆，智能机器也许可以存在于数学和物理学的虚拟世界中并进行思考。例如，在数学等理科领域里，有些问题是要求能够理解物体在多于三维的空间里是如何运动的。那些研究空间本质的弦理论家(string theorists)，认为宇宙有十维或更多的维度。人类思考四维或四维以上的数学问题就已经非常费劲，也许合理设计的智能机器能如你我理解三维空间一样去理解多维空间，并能很好地预测它们的运动规律。

最后，我们可以将许多智能系统组合成巨大的多级系统，就好比在大脑皮层的顶层将听觉、触觉和视觉整合起来。这样的系统可以通过智能机器群的方式自动学习、建立和预测模式。通过国际互联网络之类的分布式通信媒介，这些智能机器可以分布于地球的每个角落。更多的层级可以学习更深奥难懂

的模式，并能进行更复杂的类推。

做这些设想目的是为了说明类大脑的机器可以在许多方面以惊人的幅度超越我们自身的能力。它们可以以快上 100 万倍的速度思考和学习，可以记忆庞大的详细信息，还可以理解相当抽象的模式；它们有比我们敏感得多的感官，这些感官可以分散四处，能洞察细微的现象；它们还可以思考三维、四维或更多维度的问题。所有这些有趣的设想不能让智能机器模仿人类或是让复杂的机器人来实现。

现在，我们已充分认识到，把智能等同于人类行为的“图灵测试”限制了我们的想像力。首先理解了什么是智能以后，我们就可以制造出不仅仅是复制人类行为的、更有价值的智能机器，它将会成为神奇的工具，将大幅度地扩展我们对宇宙的认识。

这些设想需要多久才能实现呢？50 年后我们能造出智能机器吗？20 年怎样？抑或只需 5 年呢？在高科技领域有这么一种说法：变化比你的短期预期要长，但比你的长期预期要短。这句话曾无数次地被应验。有人在会上宣布了某项新技术，并宣称 4 年后它将普及到每个家庭。结果他错了。4 年过去了，8 年过去了，人们开始认为这根本不可能发生。就在毫无指望的时候，这项新技术开始普及，并引起了轰动。智能机器领域也会出现类似的情形，在初始阶段会进展很慢，但很快就会腾飞起来。

在神经科学会议上，我会走遍会场，询问每个人，问他们

需要多久我们才会有可行的大脑皮层理论。有那么一些人——不到 5%——会说“永远不会有！”或“我们不是已经有了吗？”（对于他们的职业来说这是多么令人不解的回答啊！）另外有 5% 的人会说 5~10 年。剩下的人中有一半会说 10~50 年，或者说“在我有生之年吧”。最后那些人会说 50~200 年，或者说“我今生是看不到了”。我支持那些对此充满信心的人。几十年来，我们都处在“缓慢”进展的时期里。因此，在许多人的看来，理论神经科学的发展和智能机器的开发已经完全停顿下来了。以过去 30 年已取得的成果为参照，我们很自然就会认为离最终的答案还十分遥远，但我相信这一领域现在正处在转折点上，很快就会腾飞。

我们可以加速未来的进程，可以使转折点快一点到来。本书的目的之一就是要让你相信，只要有了正确的理论框架，就可以在理解大脑皮层方面取得飞快的进展；要让你相信，有了记忆-预测理论框架作为指引，可以揭示大脑工作的原理以及我们是如何思考的。我们需要这些知识，有了它们才能制造出智能机器。如果这一理论模型是正确的话，整个进程将会大大加快。

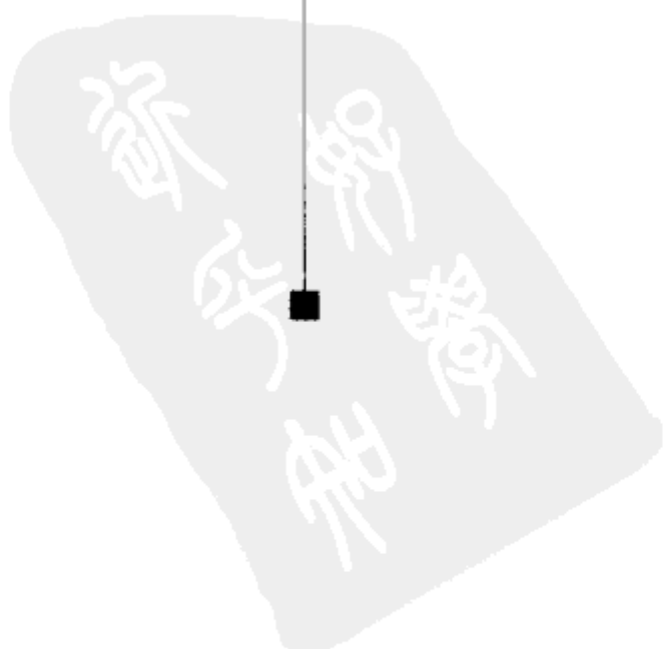
因此，虽然不愿意预言智能机器成为现实的具体时间，但我相信，只要今天有足够多的人能全力解决这一问题，那么，几年后我们就有可能建立起有用的原型和脑皮层模拟装置。我希望，在 10 年内智能机器将成为最热门的科技领域。对此，我不愿意谈得过于具体，因为我知道很容易低估重要事件发生所需要的时间。那为什么对于理解大脑和制造智能机器的进展速度我会如此乐观呢？这是因为我研究智能问题已有相当长的时间。还是在 1979 年，当我迷上大脑研究时，就觉得解决智能这

一难题是能在我有生之年可以完成的。多年来，我细心观察了人工智能和神经网络的盛衰，以及称之为“大脑的十年(the Decade of the Brain)”的 20 世纪 90 年代变化发展。我也亲眼目睹了人们对于理论生物学，尤其是理论神经科学的看法不断在转变。我也看到了有关预测的思想、层次表征和时间因素都被引入了神经科学。我还知道，我和我的同事对大脑的理解也在加深。我兴奋于 18 年前提出了预测的作用及随后对预测所作的测试。因为 20 年来我一直埋头于神经科学和计算机领域，也许我的大脑已经建立了关于科技将如何变化发展的高级模型，这一模型预测到我们即将面临快速发展，现在就是转折点。



结 语

Epilogue





宇航员卡尔·萨根(Carl Sagan)曾说过，对一件事物的理解并不会损其神奇性、减少其神秘感。许多人担心科学的理解注定以失去神奇性为代价，好像知识会使生活的滋味与色彩荡然无存。但萨根是正确的。事实是随着理解的加深，我们越来越适应自己在宇宙中的角色；同时，宇宙也越发神秘妖娆。作为一个活生生、有意识、具有智能和创造性的微小粒子生活在广袤无垠的宇宙中，远远要比生活在以扁平狭窄的地球为中心的小宇宙上要精彩得多。同样，理解了我们的大脑如何运转，也不会减少天地万物、我们的生活以及我们未来的神奇性和神秘感。当我们将对于大脑的理解运用于了解自己、制造智能机器和获取更多的知识上时，我们只会愈发惊叹不已。

当我们接受这个挑战时，我想起了物理学家埃尔文·薛丁格(Erwin Schrodinger)，他于1944年写了一本名为《生命是什么》的小册子。在这本书中，他呼吁年轻的科学家们意识到，一个生物体的活动需要精确的自然法则；此外，遗传特征，这一由染色体用某种方式编写的密码一定是可以破解的。在詹姆斯·沃森(James Watson)和弗朗西斯·克里克(Francis Crick)于1953年发现基因密码之前，薛丁格描述了突变和熵之谜。他指出有机体得以维持，是依赖于从环境中获得秩序。许多上个世纪最成功的生物学家都在中学或大学时代读过《生命是什么》，他们说这本书改变了他们的生活轨迹。

我希望我的这本书可以促使年轻的工程师和科学家去研究脑皮层、采用记忆-预测体系并制造出智能机器。人工智能在其鼎盛时期是一场规模浩大的运动，它拥有自己的期刊、学位课程、书籍、商业计划和企业。神经网络在20世纪80年代

兴盛时同样也让人们为之兴奋无比。但是人工智能和神经网络的科学体系框架并不适用于制造智能机器。

我建议我们现在沿着一条更有希望的新的道路前行。如果你正就读于中学或大学，这本书会激发你对这一技术进行研究、激发你制造最早和真正意义上的智能机器、激励你帮助开创一个新的行业，我鼓励你现在就做，并使之成为现实。企业成功的秘密之一就是在百分之百地确信自己能成功之前就投身于新的领域。把握时机也很重要，如果涉入过早，你会太费力；如果你一直等到不确定因素都不存在了，那就太迟了。我坚信现在就是开始设计和制造类似脑皮层记忆系统的时机。这一领域将在科学和商业上都无比重要。在未来的 10 年内，建立在层级体系记忆之上的这个新行业将会出现诸如英特尔和微软这样的企业。如此规模的尝试在经济上是有风险的，对智力的要求也颇高，但值得一试。我希望你能和我、和其他接受挑战的人们一起，开创一项世人闻所未闻的最伟大的技术。



附录

可检验的预言

Testable Predictions



每个理论都会引出可检验的假设，因为实证性检验是鉴定新思想正确性的唯一确定的方法。所幸的是，记忆 - 预测理论框架是以生物学为基础的，因而它所引出的是若干能被检验的具体的、新的假设。在这个附录里，我列举了其中一些假设，它们可以证伪或证实本书所提出的设想。这些内容比第六章的内容更加艰深难懂，因此，能否掌握这些内容，并不会影响大家对本书其他部分的内容的理解。在下面的这些假设当中，有几个需要在清醒的动物或人的身上进行实验，因为这些实验需要受试对象对刺激做出通常是在清醒状态下才能做出的预期或预测。以下的这些假设并不是按其重要性进行排列的。

假设一

大脑皮层的所有区域，包括基本感觉皮层区在内，都应该有能对所感觉事件产生预期并表现出强烈兴奋度的细胞，它们不仅仅只对感觉事件做出反应。比如，托尼·扎多(Tony Zador)在冷泉港实验室曾做过实验表明，小鼠的基本听觉皮层中有些细胞会在它期待听到声音而实际并没有声音的时候变得很兴奋。这应该属于脑皮层的一般属性，在视觉皮层区和体觉皮层区都应该有类似的预期性兴奋。细胞能对某种感觉输入产生预期并变得兴奋，这就是预测，这是记忆 - 预测理论模型的基本前提。

假设二

一个预测在空间上越是具体，那些因预期这一事件而变得兴奋的细胞就越靠近基本感觉皮层。

如果我们通过视觉模式序列来训练一只猴子，使它能在非常精确的时间里预期一个特定的视觉模式，那么，当它预期到所期待的模式时，一些细胞就会马上表现出强烈的兴奋。如果那只猴子学会了期待见到一张面孔，而它却不能确切知道是什么面孔以及什么时候才会出现那张面孔的话，我们就会发现那些期待性的细胞出现在面孔识别区域，而不是较低的视觉区域。然而，如果猴子注视着目标，并已经学会了期待特定的模式在视野的准确位置上出现，那么，那些期待性的细胞就应该位于 V1 区或接近 V1 区的地方。用来表示预测的细胞兴奋可以沿着皮层的层级体系一直向下传递，其传递的远近取决于预测的特性。有时，这种兴奋可以一直传到基本感觉区，有时传到较高区域的时候就会停下来。其他感觉形式也应该存在着类似的情况。

假设三

那些因期待感觉输入而强烈兴奋的细胞应该倾向于出现在皮层的第 2 层、第 3 层和第 6 层，预测应该会在体系的第 2 层和第 3 层中停止向下传播。

预测是这样沿着皮层体系向下传播的：先经由第 2 层、第 3 层传播，然后投射到第 6 层。第 6 层中的细胞再扩散到下面区域的第 1 层，从而激活第 2 层、第 3 层中的另一组细胞，等等。因此，就是在这些层级(第 2 层、第 3 层和第 6 层)中，我们可以看到因期待而引起的兴奋。记得我们说过，第 2 层、第 3 层中的兴奋细胞表现为一组潜在的兴奋垂直柱，也就是潜在的预测。第 6 层中的兴奋细胞表现为数量更少的一些垂直柱，

就是来自皮层的某个区域的兴奋预测。当预测沿着体系向下传播时，预测兴奋将最终停止在第2层、第3层。比如说，一只小鼠学会了预期两种不同音调中的一种，根据外部信号的提示，小鼠可以知道它将听到两个音调当中的一种，但它不能预测出是哪一种。这时，我们将看到期待性兴奋出现在第2层、第3层表征两种音调的垂直柱中。因为小鼠无法判断将听到的具体是哪个音调，所以在相应区域的第6层将不会有期待性兴奋。如果在另一个实验里，小鼠能预测出将听到的是哪个音调，那么我们将可能在第6层响应那个特定音调的垂直柱中看到期待性兴奋。我们不能完全排除在第4层和第5层中也出现期待性细胞，比如说，有可能在这些层级中存在着几类具有某些未知功能的细胞。因此，这个假设相对来说比较弱，但我觉得还是值得一提。

假设四

第2层和第3层中应该会有一类细胞能优先接受来自较高皮层区中第6层的细胞的输入。

记忆-预测模型中有一部分内容提到，学得的同时发生模式的序列将会产生一个临时的恒定表征，也就是我所称的“名字”。我的看法是，所谓“名字”就是大脑皮层中某个区域的第2层或第3层中以不同垂直柱方式组成的一组细胞，只要隶属于该序列的事件在发生，这组细胞就会保持兴奋状态（比如，只要听到某个曲子中的某个音符，就总会有一组细胞保持兴奋状态）。代表该序列名字这组细胞是通过大脑皮层较高区域中第6层的反馈而变得兴奋起来的。我认为，这些“名字”细胞

很可能是在第 2 层，因为它们与第 1 层非常接近。但也有可能是第 2 层和第 3 层中的任何一类细胞，它们有树突伸展到第 1 层。如果我们希望这个命名系统能正常运作的话，这些“名字”细胞顶部的树突就必须倾向于与位于第 1 层而源于较高区域中第 6 层的轴突形成突触。它们应该避免与位于第 1 层而源于丘脑的轴突形成突触。因此，我们的理论是，在第 2 层和第 3 层有一类细胞，它们顶部的树突伸到第 1 层，并倾向于与源于上面区域中第 6 层细胞的轴突形成突触。与第 1 层细胞有突触的其他细胞不会有这种倾向性。在我看来，这是一个很强的新假设。

必然地，我们会有另一个假设，那就是，在第 2 层或第 3 层中有另一类细胞，它们顶部的树突有选择性地与源于丘脑中非特定区域的轴突形成突触。这些细胞会预测序列中接下来将要出现的项目。

假设五

在假设四中，我们讲到有一组“名字”细胞在已学得序列出现期间都将一直保持兴奋。

在已学得序列出现期间一直保持兴奋的这组细胞就是一个可预测序列的“名字”。因此，即使当垂直柱中其他部分的细胞的兴奋状态在发生改变，也会有细胞组一直保持兴奋。但不幸的是，我们不知道这些“名字”细胞的兴奋状态是什么样子的。举例来说，一个名字模式的恒定兴奋状态可以非常简单，可以是这组“名字”细胞集体一致发出单尖峰脉冲信号，因此，这组兴奋细胞可能很难被侦测到。

假设六

第2层或第3层中还有另外一类细胞(与假设四和假设五中的“名字”细胞不同),它们会因响应预料外的输入而变得兴奋,但不会对已预测到的输入产生反应。

之所以要有这个假设,是因为预料外的事件必须沿着皮层体系向上传递,但是,当一个事件已经被预测到时,我们并不会原原本本地将它沿体系向上传递,因为它在本地已被预测到了。因此,第2层和第3层中要有一类细胞,它们应与假设四和假设五中提到的“名字”细胞有所不同。这类细胞会对预料外的事件产生反应并兴奋起来,而对已预测到的事件不产生反应。这些细胞的轴突应该延伸到皮层中的较高区域。我提出了一种机制,可以完成这两种状态间的变化。这类细胞可以通过由“名字”细胞激活的中间神经元而被抑制,但现在还没法提出一个针对该机制的实质性假设。我们所知道的仅仅是有一些细胞应该表现出不同的兴奋方式。在我看来,这又是一个很强的新假设。

假设七

如假设六所说,预料外的事件应该沿体系向上传递。事件越是新颖、不常见,这种预料外的输入就会传递得越高。完全新颖的事件应该到达海马。

经过大量学习的模式在体系的较低层级就得到预测,而相反,如果一个输入越是新颖,它就会被传递到体系的越高层级。我们应该有可能设计出一个实验来找出其中的差异。比如说,某人在听不熟悉但很简单的曲子,如果他听到某个音符,

虽说这个音符是未被预测到的，但它与整个音乐的风格相符，那么这个意外的音符就会引起听觉皮层的兴奋变化，这种变化会在皮层体系中向上扩散，到达较高级的层级。然而，如果听到的音符与音乐风格不相符，他听到的是完全没意义的声音，比如什么东西的破碎声，我们将会发现由此声音引起的兴奋变化是沿皮层体系传递到更高的层级。如果受试者期待听到一声破碎声，而实际听到是一个音符，那结果就会相反。这个假设可以通过功能性核磁共振成像(fMRI)技术在人脑上得到验证。

假设八

顿悟会引起准确的阶流式预测兴奋沿皮层体系向下传播。

当一个令你迷惑的感觉模式最终被理解，比如在识别图 12 中的斑点狗的过程中，你发出“啊哈”的惊叹声时，大脑皮层的某个区域正试图调用记忆来匹配输入的信息。如果在较低的区域完成了匹配，那么预测将会在皮层体系中被持续向下传递，一直到达所有的下级区域。假设这是对刺激的正确辨识，那么体系中所有的区域都会迅速确定正确的预测。当一幅图画有两种理解时，比如，一个花瓶的轮廓能被看作两副面孔或一个瓶颈(当一个物体的图像在两种不同方向上交替出现时，也会出现同样的结果)。每当对这幅图画的感知发生变化时，我们应该可以看到新的预测将沿着体系向下传播。在最底层的区域，比如说 V1 区，表征图画中一条线段的细胞垂直柱将在图画的两种感知过程中一直保持兴奋(假定眼睛没有移动)。然而，我们会看到在那个细胞垂直柱中有些细胞会处于交替兴奋状态。也就是说，同一图画中存在有相同的低级特征，但随着人们对

它的不同解读，同一垂直柱中有不同的细胞会变得兴奋。重要的是，当高级的感知发生变化时，在皮层体系中就会有不同的预测流向传播。

对已知的视觉对象的每一次扫视，都会引起类似的预测流传递。

假设九

根据记忆 - 预测模型，锥体细胞能探测到纤细树突上精确、一致的突触输入。

多年来，人们一直认为神经元很有可能是简单的信息综合单元，它是通过对来自其上所有突触的输入进行整合来决定它自己是否要发出脉冲。而今天的神经科学对神经元是如何工作的问题还有许多不清楚的地方。仍然有人认为，神经元不过是简单的信息综合单元，而且有许多神经网络模型就是基于这样的认识而建立的。还有许多关于神经元的模型认为，神经元在工作时它的树突上的各枝都彼此独立地运作。但根据记忆 - 预测模型，神经元在一个狭窄的时间窗里只需探测若干一致的兴奋突触。只需单个足以引起细胞兴奋增强的突触，这个模型就可以工作。但很可能是在纤细树突上有彼此接近的两个或更多的兴奋突触。因此，一个有着成千上万个突触的神经元就能学会对各种精确且独立的输入模式而产生兴奋。这并不是什么新的设想，目前已经有证据可以支持它，只是这种设想与多年来标准的模型有着根本性的区别。如果发现神经元并不是对精确而分散的输入模式产生兴奋的话，我们就很难维持记忆 - 预测理论的完整性。其实，那些细胞体之上或接近细胞体的粗厚树突上的突触不必以同样方式工

作，它们只需纤细树突上的突触有这样的表现就可以了。

假设十

随着我们远离基本感觉皮层区沿着皮层体系向上进行观察时，轴突和树突的伸展会越来越宽大。我认为，V1 区以及其他的基本感觉区并不是大区域，它们实际上是由许多小区域组成的。组成这些基本感觉区的次区域也许是大脑皮层中最小的区域。我们对 V1 区和其他感觉区的这种看法能帮我们理解皮层体系的工作原理，它与我们对这些基本感觉区的许多认识并不矛盾。这些小区域在生理上的表现是，这些基本感觉区中的轴突和树突的伸展应该要窄小一些，而随着我们观察到 V1 区、V2 区和 V4 区，它们将会变得宽大一些。这种扩展分布基本上能反映一个区域的大小。已经有实验证据可以支持这种看法，因此这也不是一个新的假设。然而，在视觉区、体觉区、皮层运动区甚至可能在听觉皮层区，也应该存在这种轴突和树突的扩展。

假设十一

表征将随着训练在体系中逐步向下形成。

我曾阐明过，经过反复的训练，大脑皮层在其较低的区域里会重新学得序列，对模式序列的记忆将改变传递到皮层中更高层级的输入模式，基于此，我们便可以很自然地推出以上观点。这种信息加工过程会有几种结果。其一就是，应该会存在这样的细胞，那就是经过大量的训练，它们会对皮层中较低层级中复杂的刺激产生反应；经过适当的训练，它们对皮层中较

高层级的刺激也能产生反应。比如，在人的大脑中，我们可以在 IT 区这样的区域找到这样的细胞，它们经过单个字母的识别训练后，就能对印刷字母产生反应。然而，当经过训练认读一个个完整的单词时，我们应该可以在除 IT 区以外的区域，比如 V4 区的不同部分找到能响应字母的细胞。如果我们试验其他生物体、利用其他刺激、观察其他区域的话，也应该能得到类似的结果。这一学习过程的另一结果就是，回忆和错误被检测到的地方应该会变动。也就是说，关于牢固记忆模式的感觉信息将不会在皮层体系中向上传递多远。这应该可以通过影像技术得到证实，因为输入的信息不必在皮层中传递太远就可以得到识别并回忆起来，所以，我们应该可以测定，某种刺激在反应时也会有所变化。

假设十二

恒定表征应存在于所有的皮层区域。

大家都知道存在着这样的细胞，它们有选择性地响应不随细节而变化的输入。我们都已经观察到了那些对面孔、手以及比尔·克林顿等产生响应的细胞。记忆-预测模型预言，大脑皮层的所有区域都应该形成恒定表征。这些恒定表征应反映某个皮层区域以下的所有感觉形式。比如，如果我的视觉皮层中有一个对比尔·克林顿产生反应的细胞，那么，当我见到他时这个细胞就会兴奋起来。如果我的听觉皮层中有一个对比尔·克林顿产生反应的细胞，那么，当我听到他的姓名时这个细胞就会兴奋起来。在联合区中，我们也将能找到一些接受因为看到或听到比尔·克林顿而产生的视、听觉输入的细胞。我们将

找到所有感觉形式的恒定表征，甚至还包括皮层运动区的恒定表征。在皮层运动区，细胞将表征复杂的运动序列，越是在运动区的高级层级，表征就越复杂也越恒定。(最近有研究发现，在猴子身上能激活由手到口的复杂运动的细胞。)这些不是什么新的假设。大多数的研究人员相信，恒定表征在整个大脑皮层的许多地方都有形成。然而，虽然我这么说，但还没有实验能证明的确各处都有。记忆 - 预测模型预言，我们的大脑皮层的每个区域都将能找到这样的细胞。

以上的假设是一些可以检验本书所提模型的方法。当然，肯定还会有其他的方法。然而，我们只可能证明一个理论是错误的，而不可能证明一个理论是正确的。因此，即使以上所有假设都被证明是正确的，也不能证明记忆 - 预测模型的假设就是正确的，而只能为支持该理论提供一些有力的证据。反之亦然，如果以上假设中有一些被证明是错误的，也不一定就使整个理论论断无效。对于某些假设来说，可能存在其他的方法来获得所需的行为表现。比如说，序列的名字可以通过其他途径来产生。本附录的目的仅仅是希望告诉大家，我们的理论模型可以引出若干假设，而这些假设又都是可以检验的。设计实验是具有挑战性的工作，它不是本书所应涵盖的范围。利用功能性核磁共振成像技术等方法来检验本理论也是个不错的想法。目前有许多先进的影像实验室，比起直接从脑细胞记录数据的方法，在这些影像实验室进行实验将能较快地完成检验任务。